



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :  C12N 15/54, 9/12		A1	(11) International Publication Number: WO 92/03556  (43) International Publication Date: 5 March 1992 (05.03.92)
(21) International Application Number: PCT/US91/05753 (22) International Filing Date: 13 August 1991 (13.08.91) (30) Priority data: 567,244 13 August 1990 (13.08.90) US		(72) Inventors; and (75) Inventors/Applicants (for US only) : GELFAND, David, H. [US/US]; 6208 Chelton Drive, Oakland, CA 94611 (US). LAWYER, Frances, C. [US/US]; 6641 Saroni Drive, Oakland, CA 94611 (US). STOFFEL, Susanne [US/US]; 935 Galvin Drive, El Cerrito, CA 94530 (US). (74) Agent: KASTER, Kevin, R.; Cetus Corporation, 1400 Fifty-Third Street, Emeryville, CA 94608 (US). (81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US.	
(60) Parent Application or Grant (63) Related by Continuation US 567,244 (CIP) Filed on 13 August 1990 (13.08.90)			
(71) Applicant (for all designated States except US): CETUS CORPORATION [US/US]; 1400 Fifty-Third Street, Emeryville, CA 94608 (US).			

(54) Title: PURIFIED THERMOSTABLE NUCLEIC ACID POLYMERASE ENZYME FROM *TERMOTOGA MARITIMA*

## (57) Abstract

A purified thermostable enzyme is derived from the eubacterium *Thermotoga maritima*. The enzyme has a molecular weight as determined by gel electrophoresis of about 97 kilodaltons and DNA polymerase I activity. The enzyme can be produced from native or recombinant host cells and can be used with primers and nucleoside triphosphates in a temperature-cycling chain reaction where at least one nucleic acid sequence is amplified in quantity from an existing sequence.

## Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	CR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU <sup>+</sup>	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TC	Togo
DE*	Germany	MC	Monaco	US	United States of America

+ Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

PURIFIED THERMOSTABLE NUCLEIC ACID POLYMERASE  
ENZYME FROM  THERMOTOGA MARITIMA

Technical Field

The present invention relates to a purified, thermostable DNA polymerase 5 purified from the hyperthermophilic eubacteria Thermotoga maritima and means for isolating and producing the enzyme. Thermostable DNA polymerases are useful in many recombinant DNA techniques, especially nucleic acid amplification by the polymerase chain reaction (PCR).

Background Art

10 In Huber *et al.*, 1986, *Arch. Microbiol.* 144:324-333, the isolation of the bacterium Thermotoga maritima is described. T. maritima is a eubacterium that is strictly anaerobic, rod-shaped, fermentative, hyperthermophilic, and grows between 55°C and 90°C, with an optimum growth temperature of about 80°C. This eubacterium has been isolated from geothermally heated sea floors in Italy and the Azores. T. maritima cells have a sheath-like structure and monotrichous flagellation. T. maritima is classified in the eubacterial kingdom by virtue of having murein and fatty acid-containing lipids, diphtheria-toxin-resistant elongation factor 2, an RNA polymerase subunit pattern, and sensitivity to antibiotics.

15 Extensive research has been conducted on the isolation of DNA polymerases from mesophilic microorganisms such as E. coli. See, for example, Bessman *et al.*, 1957, *J. Biol. Chem.* 223:171-177, and Buttin and Kornberg, 1966, *J. Biol. Chem.* 241:5419-5427. Much less investigation has been made on the isolation and purification of DNA polymerases from thermophiles such as Thermotoga maritima. In Kaledin *et al.*, 1980, *Biokhimiya* 45:644-651, a six-step isolation and enrichment 20 procedure for DNA polymerase activity from cells of a Thermus aquaticus YT-1 strain is disclosed. These steps involve isolation of crude extract, DEAE-cellulose chromatography, fractionation on hydroxyapatite, fractionation on DEAE-cellulose, and chromatography on single-strand DNA-cellulose. The molecular weight of the purified enzyme is reported by Kaledin *et al.* as 62,000 daltons per monomeric unit.

25 30 A second enrichment scheme for a polymerase from Thermus aquaticus is described in Chien *et al.*, 1976, *J. Bacteriol.* 127:1550-1557. In this process, the crude extract is applied to a DEAE-Sephadex column. The dialyzed pooled fractions are then subjected to treatment on a phosphocellulose column. The pooled fractions are dialyzed, and bovine serum albumin (BSA) is added to prevent loss of polymerase 35 activity. The resulting mixture is loaded on a DNA-cellulose column. The pooled material from the column is dialyzed. The molecular weight of the purified protein is reported to be about 63,000 daltons to 68,000 daltons.

The use of thermostable enzymes, such as those described in Chien *et al.* and Kaledin *et al.*, to amplify existing nucleic acid sequences in amounts that are large compared to the amount initially present is described in U.S. Patent Nos. 4,683,195; 4,683,202; and 4,965,188, which describe the PCR process, each of which is 5 incorporated herein by reference. Primers, template, nucleoside triphosphates, the appropriate buffer and reaction conditions, and polymerase are used in the PCR process, which involves denaturation of target DNA, hybridization of primers, and synthesis of complementary strands. The extension product of each primer becomes a template for the production of the desired nucleic acid sequence. The patents disclose 10 that, if the polymerase employed is a thermostable enzyme, then polymerase need not be added after every denaturation step, because heat will not destroy the polymerase activity.

U.S. Patent No. 4,889,818, European Patent Publication No. 258,017, and PCT Publication No. 89/06691, the disclosures of which are incorporated herein by 15 reference, describe the isolation and recombinant expression of an ~94 kDa thermostable DNA polymerase from Thermus aquaticus and the use of that polymerase in PCR. Although T. aquaticus DNA polymerase is especially preferred for use in PCR and other recombinant DNA techniques, there remains a need for other thermostable polymerases.

20 Accordingly, there is a desire in the art to produce a purified, thermostable DNA polymerase that may be used to improve the PCR process described above and to improve the results obtained when using a thermostable DNA polymerase in other recombinant techniques, such as DNA sequencing, nick-translation, and even reverse transcription. The present invention helps meet that need by providing recombinant 25 expression vectors and purification protocols for Thermotoga maritima DNA polymerase.

#### Disclosure of Invention

The present invention provides a purified thermostable DNA polymerase I 30 enzyme that catalyzes combination of nucleoside triphosphates to form a nucleic acid strand complementary to a nucleic acid template strand. The purified enzyme is the DNA polymerase I from Thermotoga maritima (Tma) and has a molecular weight of about 97 kilodaltons (kDa) as measured by SDS-PAGE and an inferred molecular weight, from the nucleotide sequence of the Tma DNA polymerase gene, of 102 kDa. This purified material may be used in PCR to produce a given nucleic acid sequence in 35 amounts that are large compared to the amount initially present so that the sequences can be manipulated and/or analyzed easily.

The gene encoding Tma DNA polymerase enzyme from Thermotoga maritima has also been identified, cloned, sequenced, and expressed at high level and provides

yet another means to prepare the thermostable enzyme of the present invention. In addition to the intact gene and the coding sequence for the Tma enzyme, derivatives of the coding sequence for Tma DNA polymerase are also provided.

5 The invention also encompasses a stable enzyme composition comprising a purified, thermostable Tma enzyme as described above in a buffer containing one or more non-ionic polymeric detergents.

Finally, the invention provides a method of purification for the thermostable polymerase of the invention. This method involves preparing a crude extract from Thermotoga maritima cells, adjusting the ionic strength of the crude extract so that the 10 DNA polymerase dissociates from nucleic acid in the extract, subjecting the extract to hydrophobic interaction chromatography, subjecting the extract to DNA binding protein affinity chromatography, and subjecting the extract to cation or anion exchange or hydroxyapatite chromatography. In a preferred embodiment, these steps are performed sequentially in the order given above. The nucleotide binding protein affinity 15 chromatography step is preferred for separating the DNA polymerase from endonuclease proteins.

#### Modes for Carrying out the Invention

The present invention provides DNA sequences and expression vectors that encode Tma DNA polymerase, purification protocols for Tma DNA polymerase, 20 preparations of purified Tma DNA polymerase, and methods for using Tma DNA polymerase. To facilitate understanding of the invention, a number of terms are defined below.

The terms "cell," "cell line," and "cell culture" are used interchangeably and all such designations include progeny. Thus, the words "transformants" or "transformed 25 cells" include the primary transformed cell and cultures derived from that cell without regard to the number of transfers. All progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. Mutant progeny that have the same functionality as screened for in the originally transformed cell are included in the definition of transformants.

30 The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, a ribosome binding site, and possibly other sequences, such as transcription termination sequences. Eucaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

35 The term "expression system" refers to DNA sequences containing a desired coding sequence and control sequences in operable linkage, so that hosts transformed with these sequences are capable of producing the encoded proteins. To effect

transformation, the expression system may be included on a vector; the relevant DNA can also be integrated into the host chromosome.

The term "gene" refers to a DNA sequence that codes for the expression of a recoverable bioactive polypeptide or precursor. Thus, the Tma DNA polymerase gene 5 includes the promoter and Tma DNA polymerase coding sequence. The polypeptide can be encoded by a full-length coding sequence or by any portion of the coding sequence so long as the desired enzymatic activity is retained.

The term "operably linked" refers to the positioning of the coding sequence such that control sequences will function to drive expression of the encoded protein. 10 Thus, a coding sequence "operably linked" to a control sequence refers to a configuration wherein the coding sequence can be expressed under the direction of the control sequence.

The term "mixture" as it relates to mixtures containing Tma polymerase refers to a collection of materials that includes Tma polymerase but can also include other 15 proteins. If the Tma polymerase is derived from recombinant host cells, the other proteins will ordinarily be those associated with the host. Where the host is bacterial, the contaminating proteins will be bacterial proteins.

The term "non-ionic polymeric detergents" refers to surface-active agents that have no ionic charge and that are characterized, for purposes of this invention, by an 20 ability to stabilize the Tma enzyme at a pH range of from about 3.5 to about 9.5, preferably from 4 to 8.5. Numerous examples of suitable non-ionic polymeric detergents are presented in copending U.S. patent application Serial No. 387,003, filed July 28, 1989, the disclosure of which is incorporated herein by reference.

The term "oligonucleotide" as used herein is defined as a molecule comprised of 25 two or more deoxyribonucleotides or ribonucleotides, preferably more than three, and usually more than ten. The exact size will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. The oligonucleotide may be derived synthetically or by cloning.

The term "primer" as used herein refers to an oligonucleotide that is capable of 30 acting as a point of initiation of synthesis when placed under conditions in which primer extension is initiated. An oligonucleotide "primer" may occur naturally, as in a purified restriction digest, or be produced synthetically. Synthesis of a primer extension product that is complementary to a nucleic acid strand is initiated in the presence of four different nucleoside triphosphates and the Tma thermostable enzyme in 35 an appropriate buffer at a suitable temperature. A "buffer" includes cofactors (such as divalent metal ions) and salt (to provide the appropriate ionic strength), adjusted to the desired pH. For Tma polymerase, the buffer preferably contains 1 to 3 mM of a magnesium salt, preferably MgCl<sub>2</sub>, 50 to 200 μM of each nucleoside triphosphate, and 0.2 to 1 μM of each primer, along with 50 mM KCl, 10 mM Tris buffer (pH 8.0-8.4),

and 100  $\mu$ g/ml gelatin (although gelatin is not required and should be avoided in some applications, such as DNA sequencing).

The primer is single-stranded for maximum efficiency in amplification, but may alternatively be double-stranded. If double-stranded, the primer is first treated to separate its strands before being used to prepare extension products. The primer is usually an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the polymerase enzyme. The exact length of a primer will depend on many factors, such as source of primer and result desired, and the reaction temperature must be adjusted depending on primer length to ensure proper annealing of primer to template. Depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15 to 35 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable complexes with template.

A primer is selected to be "substantially" complementary to a strand of specific sequence of the template. A primer must be sufficiently complementary to hybridize with a template strand for primer elongation to occur. A primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being substantially complementary to the strand. Non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the template to hybridize and thereby form a template/primer complex for synthesis of the extension product of the primer.

The terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes that cut double-stranded DNA at or near a specific nucleotide sequence.

The term "thermostable enzyme" refers to an enzyme which is stable to heat and is heat resistant and catalyzes (facilitates) combination of the nucleotides in the proper manner to form primer extension products that are complementary to a nucleic acid strand. Generally, synthesis of a primer extension product begins at the 3' end of the primer and proceeds towards the 5' end of the template strand until synthesis terminates. A thermostable enzyme must be able to renature and regain activity after brief (i.e., 5 to 30 seconds) exposure to temperatures of 80°C to 105°C and must have a temperature optimum of above 60°C.

The Tma thermostable DNA polymerase enzyme of the present invention satisfies the requirements for effective use in the amplification reaction known as the polymerase chain reaction or PCR. The Tma DNA polymerase enzyme does not become irreversibly denatured (inactivated) when subjected to the elevated temperatures for the time necessary to effect denaturation of double-stranded nucleic acids, a key step

in the PCR process. Irreversible denaturation of an enzyme for purposes herein refers to permanent and complete loss of enzymatic activity.

The heating conditions necessary to effect nucleic acid denaturation will depend, e.g., on the buffer salt concentration and the composition, length, and amount of the nucleic acids being denatured, but typically the denaturation temperature ranges from about 80°C to about 105°C for a few seconds to minutes. Higher temperatures may be required for nucleic acid denaturation as the buffer salt concentration and/or GC composition of the nucleic acid is increased. The Tma enzyme does not become irreversibly denatured upon relatively short exposures to temperatures of about 80°C-105°C.

The Tma thermostable enzyme has an optimum temperature at which it functions that is higher than about 60°C. Temperatures below 60°C facilitate hybridization of primer to template, but depending on salt composition and concentration and primer composition and length, hybridization of primer to template can occur at higher temperatures (e.g., 60°C-80°C), which may promote specificity of the primer elongation reaction. The higher the temperature optimum for the enzyme, the greater the specificity and/or selectivity of the primer-directed extension process. The Tma enzyme exhibits activity over a broad temperature range from about 45°C to 90°C; a preferred optimum temperature is 75°C-80°C.

The present invention also provides DNA sequences encoding the thermostable DNA polymerase I activity of Thermotoga maritima. The amino acid sequence encoded by this sequence has homology to portions of the thermostable DNA polymerases of Thermus aquaticus and Thermus thermophilus. The complete coding sequence, from the 5'-ATG start codon to the TGA-3' stop codon, of the Tma DNA polymerase gene is depicted below and listed as SEQ ID NO: 1 in the sequence listing section. The sequence is numbered for reference.

```
1 ATGGCGAGAC TATTCTCTT TGATGGAAC GCTCTGGCCT ACAGAGCGTA
 51 CTATGCGCTC GATAGATCGC TTTCTACTTC CACCGGCATT CCCACAAACG
 101 CCACATACGG TGTGGCGAGG ATGCTGGTGA GATTCACTAA AGACCATATC
 151 ATTGTCGGAA AAGACTACGT TGCTGTGGCT TTGACAAAAA AAGCTGCCAC
 201 CTTCAGACAC AAGCTCCTCG AGACTTACAA GGCTCAAAGA CCAAAGACTC
 251 CGGATCTCCT GATTCAAGCAG CTTCCGTACA TAAAGAAGCT GGTCGAAGCC
 301 CTTGGAATGA AAGTGCTGGA GGTAGAAGGA TACGAAGCGG ACGATATAAT
 351 TGCCACTCTG GCTGTGAAGG GGCTTCCGCT TTTTGATGAA ATATTCAAG
 35 401 TGACCGGAGA TAAAGACATG CTTCAAGCTTG TGAACGAAAAA GATCAAGGTG
 451 TGGCGAATCG TAAAGGGAT ATCCGATCTG GAACTTACG ATGCGCAGAA
 501 GGTGAAGGAA AAATACGGTG TTGAACCCCA GCAGATCCCG GATCTTCTGG
 551 CTCTAACCGG AGATGAAATA GACAACATCC CCGGTGTAAC TGGGATAGGT
 601 GAAAAGACTG CTGTTCAGCT TCTAGAGAAG TACAAAGACC TCGAAGACAT
```

651 ACTGAATCAT GTTCGCGAAC TTCCCTCAAAA GGTGAGAAAA GCCCTGCTTC  
701 GAGACAGAGA AAACGCCATT CTCAGCAAAA AGCTGGCGAT TCTGGAAACA  
751 AACGTTCCCA TTGAAATAAA CTGGGAAGAA CTTCGCTACC AGGGCTACGA  
801 CAGAGAGAAA CTCTTACAC TTTTGAAGA ACTGGAATT GCATCCATCA  
5 851 TGAAGGAACT TCAACTGTAC GAAGAGTCCG AACCCGTTGG ATACAGAATA  
901 GTGAAAGACC TAGTGGATT TGAAAAACTC ATAGAGAAAC TGAGAGAATC  
951 CCCTTCGTTGCCATAGATC TTGAGACGTC TTCCCTCGAT CCTTCGACT  
1001 GCGACATTGT CGGTATCTCT GTGTCTTCA AACCAAAGGA AGCGTACTAC  
1051 ATACCACTCC ATCATAGAAA CGCCCAGAAC CTGGACGAAA AAGAGGTTCT  
10 1101 GAAAAAGCTC AAAGAAATTCTGGAGGACCC CGGAGCAAAG ATCGTTGGTC  
1151 AGAATTGAA ATTGATTAC AAGGTGTTGA TGGTGAAGGG TGTTGAACCT  
1201 GTTCCTCCTT ACTTCGACAC GATGATAGCG GCTTACCTTC TTGAGCCGAA  
1251 CGAAAAGAAG TTCAATCTGG ACGATCTCGC ATTGAAATTT CTTGGATACA  
1301 AAATGACATC TTACCAAGAG CTCATGTCCT TCTCTTTCC GCTGTTGGT  
15 1351 TTCAGTTTG CCGATGTTCC TGTAGAAAAA GCAGCGAACT ACTCCTGTGA  
1401 AGATGCAGAC ATCACCTACA GACTTACAA GACCCCTGAGC TTAAAACCTCC  
1451 ACGAGGCAGA TCTGGAAAAC GTGTTCTACA AGATAGAAAT GCCCCTTGTG  
1501 AACGTGCTTG CACGGATGGA ACTGAACGGT GTGTATGTGG ACACAGAGTT  
1551 CCTGAAGAAA CTCTCAGAAG AGTACGGAAA AAAACTCGAA GAACTGGCAG  
20 1601 AGGAAATATA CAGGATAGCT GGAGAGCCGT TCAACATAAA CTCACCGAAG  
1651 CAGGTTCAA GGATCCTTT TGAAAAACTC GGCATAAAAC CACGTGGTAA  
1701 AACGACGAAA ACGGGAGACT ATTCAACACG CATAGAAGTC CTCGAGGAAC  
1751 TTGCCGGTGA ACACGAAATC ATTCCCTCTGA TTCTTGAATA CAGAAAGATA  
1801 CAGAAATTGA AATCAACCTA CATAGACGCT CTTCCCAAGA TGGTCAACCC  
25 1851 AAAGACCGGA AGGATTCATG CTTCTTCAA TCAAACGGGG ACTGCCACTG  
1901 GAAGACTTAG CAGCAGCGAT CCCAATCTTC AGAACCTCCC GACGAAAAGT  
1951 GAAGAGGGAA AAGAAATCAG GAAAGCGATA GTTCCTCAGG ATCCAAACTG  
2001 GTGGATCGTC AGTGCCACT ACTCCCAAT AGAACTGAGG ATCCTCGCCC  
2051 ATCTCAGTGG TGATGAGAAT CTTTGAGGG CATTGAAAGA GGGCATCGAC  
30 2101 GTCCACACTC TACAGCTTC CAGAATATTC AACGTGAAAC CCGAAGAAGT  
2151 AACCGAAGAA ATGCCCGCG CTGGTAAAAT GTTAATTT TCCATCATAT  
2201 ACGGTGTAAC ACCTTACGGT CTGTCTGTGA GGCTTGGAGT ACCTGTGAAA  
2251 GAAGCAGAAA AGATGATCGT CAACTACTTC GTCCTCTACC CAAAGGTGCG  
2301 CGATTACATT CAGAGGGTCG TATCGGAAGC GAAAGAAAAA GGCTATGTTA  
35 2351 GAACGCTGTT TGGAAAGAAA AGAGACATAC CACAGCTCAT GGCCCGGGAC  
2401 AGGAACACAC AGGCTGAAGG AGAACGAATT GCCATAAACCA CTCCCATACA  
2451 GGGTACAGCA GCGGATATAA TAAAGCTGGC TATGATAGAA ATAGACAGGG  
2501 AACTGAAAGA AAGAAAAATG AGATCGAAGA TGATCATACA GGTCCACGAC  
2551 GAACTGGTTT TTGAAAGTGCC CAATGAGGAA AAGGACGCGC TCGTCGAGCT

2601 GGTGAAAGAC AGAATGACGA ATGTGGTAAA GCTTCAGTG CCGCTCGAAG

2651 TGGATGTAAC CATCGGCAAA ACATGGTCGT GA

Both the complete coding sequence of the Tma DNA polymerase gene and the encoded amino acid sequence in three letter abbreviation are provided in the Sequence Listing section as SEQ ID NO: 1. For convenience, the amino acid sequence encoded by the Tma DNA polymerase gene sequence is also depicted below in one letter abbreviation from amino-terminus to carboxy-terminus; the sequence is numbered for reference.

1 MARLFLFDGT ALAYRAYYAL DRSLSTSTGI PTNATYGVAR MLVRFIKDHI  
 51 IVGKDYVAVA FDKKAATFRH KLLETYKAQR PKTPDLLIQQ LPYIKKLVEA  
 10 101 LGMKVLEVEG YEADDIIATL AVKGLPLFDE IFIVTGDKDM LQLVNEKIKV  
 151 WRIVKGISDL ELYDAQKVKE KYGVEPQQIP DLLALTGDEI DNIPGVTGIG  
 201 EKTAVQLLEK YKDLEDILNH VRELPQKVRK ALLRDRENAI LSKKLAILET  
 251 NVPIEINWEE LRYQGYDREK LLPLLKELEF ASIMKELQLY ESEEPVGYRI  
 301 VKDLVEFEKL IEKLRESPSF AIDLETSSLD PFDCDIVGIS VSFKPKEAYY  
 15 351 IPLHHRNAQN LDEKEVLKKL KEILEDPGAK IVGQNLKFDY KVLMVKGVEP  
 401 VPPYFDTMIA AYLLEPNEKK FNLDLALKF LGYKMTSYQE LMSFSFPLFG  
 451 FSFADVPVEK AANYSCEDAD ITYRLYKTLS LKLHEADLEN VFYKIEMPLV  
 501 NVLARMELNG VYVDTEFLKK LSEEYGKKLE ELAEEIYRIA GEPFNINSPK  
 551 QVSRLIFEKL GIKPRGKTTK TGDYSTRIEV LEELAGEHEI IPLILEYRKI  
 20 601 QKLKSTYIDA LPKMVNPKTG RIHASFNQTG TATGRLSSSD PNQLNLPTKS  
 651 EEGKEIRKAI VPQDPNWWIV SALYSQIELR ILAHLSGDEN LLRAFEEGID  
 701 VHTLTASRIF NVKPEEVTEE MRRAGKMVN SIIYGVTPYG LSVRLGVPVK  
 751 EAEKMIVNYF VLYPKVRDYI QRVVSEAKEK GYVRTLFGRK RDIPQLMARD  
 801 RNTQAEGERI AINTPIQGTA ADIIKLAMIE IDRELKERKM RSKMIIQVHD  
 25 851 ELVFEVPNEE KDALVELVKD RMTNVVKLSV PLEVDVTIGK TWS

The one letter abbreviations for the amino acids are shown below for convenience.

F	=	Phenylalanine	H	=	Histidine
L	=	Leucine	Q	=	Glutamine
I	=	Isoleucine	N	=	Asparagine
30 M	=	Methionine	K	=	Lysine
V	=	Valine	D	=	Aspartic Acid
S	=	Serine	E	=	Glutamic Acid
P	=	Proline	C	=	Cysteine
T	=	Threonine	W	=	Tryptophan
35 A	=	Alanine	R	=	Arginine
Y	=	Tyrosine	G	=	Glycine

The coding sequence for Tma DNA polymerase I was identified by a "degenerate primer" method that has broad utility and is an important aspect of the present invention. In the degenerate primer method, DNA fragments of any

thermostable polymerase coding sequence corresponding to conserved domains of known thermostable DNA polymerases can be identified.

In one embodiment of the degenerate primer method, the corresponding conserved domains are from the coding sequences for and amino acid sequences of the thermostable DNA polymerases of Taq, Tma, and Tth. The degenerate primer method was developed by comparing the amino acid sequences of DNA polymerase I proteins from Taq, Tth, T7, and E. coli in which various conserved regions were identified. Primers corresponding to these conserved regions were then designed. As a result of the present invention, Tma sequences can be used to design other degenerate primers.

10 The generic utility of the degenerate primer process is exemplified herein by specific reference to the method as applied to cloning the Tma gene.

To clone the Tma DNA polymerase I gene, the conserved amino acid sequences were converted to all of the possible codons for each of the amino acids. Due to the degenerate nature of the genetic code, a given amino acid may be represented by several 15 different codons. Where more than one base can be present in a codon for a given amino acid, the sequence is said to be degenerate.

The primers were then synthesized as a pool of all of the possible DNA sequences that could code for a given amino acid sequence. The amount of degeneracy of a given primer pool can be determined by multiplying the number of possible 20 nucleotides at each position.

The more degenerate a primer pool, (i.e., the greater the number of individual unique primer DNA sequences within the pool), the greater the probability that one of the unique primer sequences will bind to regions of the target chromoso: 21 DNA other than the one desired -- hence, the lesser the specificity of the resulting amplification.

25 To increase the specificity of the amplification using the degenerate primers, the pools are synthesized as subsets such that the entire group of subsets includes all possible DNA sequences encoding the given amino acid sequence, but each individual subset only includes a portion: for example, one pool may contain either a G or C at a particular position while the other contains either an A or T at the same position. Each 30 of these subpools is designated with a DG number.

Both forward primers (directed from the 5' region toward the 3' region of the gene, complementary to the noncoding strand) and reverse primers (directed from the 3' region toward the 5' region of the gene, complementary to the coding strand) were designed for most of these conserved regions to clone Tma polymerase. The primers 35 were designed with restriction sites at the 5' ends to facilitate cloning. The forward primers contained a BglII restriction site (AGATCT), while the reverse primers contained an EcoRI restriction site (GAATTC). In addition, the primers contained 2 nucleotides at the 5' end to increase the efficiency of cutting at the restriction site.

Degenerate primers were then used in PCR processes in which the target nucleic acid was chromosomal DNA from Thermotoga maritima. The products of the PCR processes using a combination of forward and reverse primer pools in conjunction with a series of temperature profiles were compared. When specific products of similar size to the product generated using Taq chromosomal DNA were produced, the PCR 5 fragments were gel purified, reamplified and cloned into the vector BSM13H3:BglII (a derivative of the Stratagene vector pBSM+ in which the HindIII site of pBSM+ was converted to a BglII site). Sequences were identified as potential thermostable DNA polymerase coding sequences if the sequences were found to encode amino acid 10 sequences homologous to other known amino acid sequences in polymerase proteins, particularly those of Taq polymerase and Tth polymerase.

The portions of the Tma DNA polymerase gene were then identified in the chromosomal DNA of Thermotoga maritima by Southern blot analysis. The Tma chromosomal DNA was digested with a variety of enzymes and transferred to 15 nitrocellulose filters. Probes labeled with <sup>32</sup>P or biotin-dUTP were generated for various regions of the gene from the cloned PCR products. The probes were hybridized to the nitrocellulose-bound genomic DNA, allowing identification of the size of the chromosomal DNA fragment hybridizing to the probe. The use of probes covering the 5' and 3' regions of the gene ensures that the DNA fragment(s) contain 20 most if not all of the structural gene for the polymerase. Restriction enzymes are identified that can be used to produce fragments that contain the structural gene in a single DNA fragment or in several DNA fragments to facilitate cloning.

Once identified, the chromosomal DNA fragments encoding the Tma DNA 25 polymerase gene were cloned. Chromosomal DNA was digested with the identified restriction enzyme and size fractionated. Fractions containing the desired size range were concentrated, desalted, and cloned into the BSM13H3:BglII cloning vector. Clones were identified by hybridization using labeled probes generated from the previous cloned PCR products. The PCR products were then analyzed on polyacrylamide gels.

30 The DNA sequence and amino acid sequence shown above and the DNA compounds that encode those sequences can be used to design and construct recombinant DNA expression vectors to drive expression of Tma DNA polymerase activity in a wide variety of host cells. A DNA compound encoding all or part of the DNA sequence shown above can also be used as a probe to identify thermostable 35 polymerase-encoding DNA from other organisms, and the amino acid sequence shown above can be used to design peptides for use as immunogens to prepare antibodies that can be used to identify and purify a thermostable polymerase.

Whether produced by recombinant vectors that encode the above amino acid sequence or by native Thermotoga maritima cells, however, Tma DNA polymerase will

typically be purified prior to use in a recombinant DNA technique. The present invention provides such purification methodology.

For recovering the native protein, the cells are grown using any suitable technique. Briefly, the cells are grown in "MMS"-medium containing (per liter): NaCl (6.93 g); MgSO<sub>4</sub>·7H<sub>2</sub>O (1.75 g); MgCl<sub>2</sub>·6H<sub>2</sub>O (1.38 g); KCl (0.16 g); NaBr (25 mg); H<sub>3</sub>BO<sub>3</sub> (7.5 mg); SrCl<sub>2</sub>·6H<sub>2</sub>O (3.8 mg); KI (0.025 mg); CaCl<sub>2</sub> (0.38 g); KH<sub>2</sub>PO<sub>4</sub> (0.5 g); Na<sub>2</sub>S (0.5 g); (NH<sub>4</sub>)<sub>2</sub>Ni(SO<sub>4</sub>)<sub>2</sub> (2 mg); trace minerals (Balch et al., 1979, *Microbiol. Rev.* 43:260-296) (15 ml); resazurin (1 mg); and starch (5 g) at a pH of 6.5 (adjusted with H<sub>2</sub>SO<sub>4</sub>). For growth on solid medium, 0.8% agar (Oxoid) may be added to the medium. Reasonable growth of the cells also occurs in "SME"-medium (Stetter et al., 1983, *Syst. Appl. Microbiol.* 4:535-551) supplemented with 0.5% yeast extract, or in marine broth (Difco 2216).

After cell growth, the isolation and purification of the enzyme takes place in six stages, each of which is carried out at a temperature below room temperature, 15 preferably about 0°C to about 4°C, unless stated otherwise. In the first stage or step, the cells, if frozen, are thawed, lysed in an Aminco french pressure cell (8-20,000 psi), suspended in a buffer at about pH 7.5, and sonicated to reduce viscosity.

In the second stage, ammonium sulfate is added to the lysate to prevent the Tma DNA polymerase from binding to DNA or other cell lysate proteins. Also in the second 20 stage, Polymin P (polyethyleneimine, PEI) is added to the lysate to precipitate nucleic acids, and the lysate is centrifuged.

In the third step, ammonium sulfate is added to the supernatant, and the supernatant is loaded onto a phenyl sepharose column equilibrated with a buffer composed of TE (50 mM Tris-Cl, pH 7.5, and 1 mM EDTA) containing 0.3 M 25 ammonium sulfate and 0.5 mM DTT (dithiothreitol). The column is then washed first with the same buffer, second with TE-DTT (without ammonium sulfate), third with ethylene glycol-TE-DTT, and finally with 2 M urea in TE-DTT containing ethylene glycol. Unless the capacity of the phenylsepharose is exceeded (i.e. by loading more than ~20-30 mg of protein per ml of resin) all of the Tma polymerase activity is retained 30 by the column and elutes with the 2 M urea in TE-DTT containing ethylene glycol.

In the fourth stage, the urea eluate is applied to a heparin sepharose column which is equilibrated with 0.08 M KCl, 50 mM Tris-Cl (pH 7.5), 0.1 mM EDTA, 0.2% Tween 20 and 0.5 mM DTT. The column is then washed in the same buffer and the enzyme eluted with a linear gradient of 0.08 M to 0.5 M KCl buffer. The peak 35 activity fractions were found at 0.225 M to 0.275 M KCl.

In the fifth stage, the fraction collected in the fourth stage is diluted with affigel-blue buffer without KCl and applied to an affigel-blue column equilibrated in 25 mM Tris-Cl (pH 7.5), 0.1 mM EDTA, 0.2% Tween 20, 0.5 mM DTT, and 0.15 M KCl. The column is washed with the same buffer and eluted with a linear gradient of 0.15 M

to 0.7 M KCl in the same buffer. The peak activity fractions were found at the 0.3 M to 0.55 M KCl section of the gradient. These fractions of peak activity are then tested for contaminating deoxyribonucleases (endonucleases and exonucleases) using any suitable procedure. As an example, endonuclease activity may be determined 5 electrophoretically from the change in molecular weight of phage  $\lambda$  DNA or supercoiled plasmid DNA after incubation with an excess of DNA polymerase. Similarly, exonuclease activity may be determined electrophoretically from the change in molecular weight of restriction enzyme digested DNA after incubation with an excess of DNA polymerase. The fractions that have no deoxyribonuclease activity are pooled and 10 diafiltered into phosphocellulose buffer containing 50 mM KCl.

Finally, in a sixth stage, the diafiltered pool from stage five is loaded onto a phosphocellulose column equilibrated to the correct pH and ionic strength of 25 mM Tris-Cl (pH 7.5), 50 mM KCl, 0.1 mM EDTA, 0.2% Tween 20, and 0.5 mM DTT. The column is then washed with the same buffer and eluted with a linear 0.05 M to 0.5 15 M KCl gradient. The peak fractions eluted between 0.215 M and 0.31 M KCl. An undegraded, purified DNA polymerase from these fractions is evidenced by an unchanged migration pattern in an *in situ* activity gel.

The molecular weight of the DNA polymerase purified from Thermotoga maritima may be determined by any technique, for example, by SDS-PAGE analysis 20 using protein molecular weight markers or by calculation from the coding sequence. The molecular weight of the DNA polymerase purified from Thermotoga maritima is determined by SDS-PAGE to be about 97 kDa. Based on the predicted amino acid sequence, the molecular weight is estimated at about 102 kDa. The purification protocol of native Tma DNA polymerase is described in more detail in Example 1. 25 Purification of the recombinant Tma polymerase of the invention can be carried out with similar methodology.

Biologically active recombinant Tma polymerases of various molecular weights can be prepared by the methods and vectors of the present invention. Even when the 30 complete coding sequence of the Tma DNA polymerase gene is present in an expression vector in E. coli, the cells produce a truncated polymerase, formed by translation starting with the methionine codon at position 140. One can also use recombinant means to produce a truncated polymerase corresponding to the protein produced by initiating translation at the methionine codon at position 284 of the Tma coding sequence. The polymerase lacking amino acids 1 through 139 (about 86 kDa), 35 and the polymerase lacking amino acids 1 through 283 (about 70 kDa) of the wild type Tma polymerase retain polymerase activity but have attenuated 5'  $\rightarrow$  3' exonuclease activity. In addition, the 70 kDa polymerase is significantly more thermostable than native Tma polymerase.

Thus, the entire sequence of the intact Tma DNA polymerase I enzyme is not required for activity. Portions of the Tma DNA polymerase I coding sequence can be used in recombinant DNA techniques to produce a biologically active gene product with DNA polymerase activity. The availability of DNA encoding the Tma DNA polymerase sequence provides the opportunity to modify the coding sequence so as to generate 5 mutagen (mutant protein) forms also having DNA polymerase activity. The amino(N)- terminal portion of the Tma polymerase is not necessary for polymerase activity but rather encodes the 5'→3' exonuclease activity of the protein. Using recombinant DNA methodology, one can delete approximately up to one-third of the N-terminal coding 10 sequence of the Tma gene, clone, and express a gene product that is quite active in polymerase assays but, depending on the extent of the deletion, has no 5'→3' exonuclease activity. Because certain N-terminal shortened forms of the polymerase are active, the gene constructs used for expression of these polymerases can include the corresponding shortened forms of the coding sequence.

15 In addition to the N-terminal deletions, individual amino acid residues in the peptide chain of Tma polymerase may be modified by oxidation, reduction, or other derivation, and the protein may be cleaved to obtain fragments that retain activity. Such alterations that do not destroy activity do not remove the protein from the definition of a protein with Tma polymerase activity and so are specifically included within the scope 20 of the present invention.

Modifications to the primary structure of the Tma DNA polymerase coding sequence by deletion, addition, or alteration so as to change the amino acids incorporated into the Tma DNA polymerase during translation of the mRNA produced from that coding sequence can be made without destroying the high temperature DNA 25 polymerase activity of the protein. Such substitutions or other alterations result in the production of proteins having an amino acid sequence encoded by DNA falling within the contemplated scope of the present invention. Likewise, the cloned genomic sequence, or homologous synthetic sequences, of the Tma DNA polymerase gene can be used to express a fusion polypeptide with Tma DNA polymerase activity or to 30 express a protein with an amino acid sequence identical to that of native Tma DNA polymerase. In addition, such expression can be directed by the Tma DNA polymerase gene control sequences or by a control sequence that functions in whatever host is chosen to express the Tma DNA polymerase.

35 Thus, the present invention provides a coding sequence for Tma DNA polymerase from which expression vectors applicable to a variety of host systems can be constructed and the coding sequence expressed. Portions of the Tma polymerase-encoding sequence are also useful as probes to retrieve other thermostable polymerase-encoding sequences in a variety of species. Accordingly, oligonucleotide probes that encode at least four to six amino acids can be synthesized and used to retrieve additional

DNAs encoding a thermostable polymerase. Because there may not be an exact match between the nucleotide sequence of the thermostable DNA polymerase gene of Thermotoga maritima and the corresponding gene of other species, oligomers containing approximately 12-18 nucleotides (encoding the four to six amino sequence) 5 are usually necessary to obtain hybridization under conditions of sufficient stringency to eliminate false positives. Sequences encoding six amino acids supply ample information for such probes. Such oligonucleotide probes can be used as primers in the degenerate priming method of the invention to obtain thermostable polymerase encoding sequences.

10 The present invention, by providing coding sequences and amino acid sequences for Tma DNA polymerase, therefore enables the isolation of other thermostable polymerase enzymes and the coding sequences for those enzymes. The amino acid sequence of the Tma DNA polymerase I protein is very similar to the amino acid sequences for the thermostable DNA polymerases of Taq and Tth. These 15 similarities facilitated the identification and isolation of the Tma DNA polymerase coding sequence. The areas of similarity in the coding sequences of these three thermostable DNA polymerases can be readily observed by aligning the sequences.

However, regions of dissimilarity between the coding sequences of the three thermostable DNA polymerases can also be used as probes to identify other 20 thermostable polymerase coding sequences that encode thermostable polymerase enzymes. For example, the coding sequence for a thermostable polymerase having some properties of Taq and other divergent properties of Tma may be identified by using probes directed to sequences that encode the regions of dissimilarity between Taq and Tma. Specifically, such regions include a stretch of four or more contiguous 25 amino acids from any one or more of the following regions, identified by amino acid sequence coordinates (numbering is inclusive): 5-10, 73-79, 113-119, 134-145, 191-196, 328-340, 348-352, 382-387, 405-414, 467-470, 495-499, 506-512, 555-559, 579-584, 595-599, 650-655, 732-742, 820-825, 850-856. These regions 30 may be considered as "hallmark motifs" and define additional regions of critical amino acid signature sequences for thermostable DNA polymerase functions (e.g. 5'→3' exonuclease activity, 3'→5' exonuclease activity, and DNA polymerase activity).

One property found in the Tma DNA polymerase, but lacking in native Taq DNA polymerase and native Tth DNA polymerase, is 3'→5' exonuclease activity. This 3'→5' exonuclease activity is generally considered to be desirable, because 35 misincorporated or unmatched bases of the synthesized nucleic acid sequence are eliminated by this activity. Therefore, the fidelity of PCR utilizing a polymerase with 3'→5' exonuclease activity (e.g. Tma DNA polymerase) is increased. The 3'→5' exonuclease activity found in Tma DNA polymerase also decreases the probability of the formation of primer/dimer complexes in PCR. The 3'→5' exonuclease activity in

effect prevents any extra dNTPs from attaching to the 3' end of the primer in a non-template dependent fashion by removing any nucleotide that is attached in a non-template dependent fashion. The 3'→5' exonuclease activity can eliminate single-stranded DNAs, such as primers or single-stranded template. In essence, every 3'-nucleotide of a single-stranded primer or template is treated by the enzyme as unmatched and is therefore degraded. To avoid primer degradation in PCR, one can add phosphorothioate to the 3' ends of the primers. Phosphorothioate modified nucleotides are more resistant to removal by 3'→5' exonucleases.

A "motif" or characteristic "signature sequence" of amino acids critical for 3'→5' exonuclease activity in thermostable DNA polymerases can be defined as comprising three short domains. Below, these domains are identified as A, B, and C, with critical amino acid residues shown in one letter abbreviation and non-critical residues identified as "x."

15	<u>Domain</u>	<u>Sequence</u>	Representative
			<u>Tma Coordinates</u>
	A	DxExxxL	323-329
	B	NxxxDxxxL	385-393
	C	YxxxD	464-468

The distance between region A and region B is 55-65 amino acids. The distance between region B and region C is 67-75 amino acids, preferably about 70 amino acids. In Tma DNA polymerase, the amino acids that do not define the critical motif signature sequence amino acids are L and TSS, respectively, in domain A; LKF and YKV, respectively, in domain B; and SCE in domain C. Domain A is therefore DLETSSL; domain B is NLKFDYKVL; and domain C is YSCED in Tma DNA polymerase I. Thus, the present invention provides a thermostable DNA polymerase possessing 3'→5' exonuclease activity that comprises domains A, B, and C, and, more particularly comprises the sequence D-X-E-X<sup>3</sup>-L-X<sup>55-65</sup>-N-X<sup>3</sup>-D-X<sup>3</sup>-L-X<sup>65-75</sup>-Y-X<sup>3</sup>-D, where one letter amino acid abbreviation is used, and X<sup>N</sup> represents the number (N) of non-critical amino acids between the specified amino acids.

A thermostable 3'→5' exonuclease domain is represented by amino acids 291 through 484 of Tma DNA polymerase. Accordingly, "domain shuffling" or construction of "thermostable chimeric DNA polymerases" may be used to provide thermostable DNA polymerases containing novel properties. For example, substitution of the Tma DNA polymerase coding sequence comprising codons about 291 through about 484 for the Thermus aquaticus DNA polymerase I codons 289-422 would yield a novel thermostable DNA polymerase containing the 5'→3' exonuclease domain of Taq DNA polymerase (1-289), the 3'→5' exonuclease domain of Tma DNA polymerase (291-484), and the DNA polymerase domain of Taq DNA polymerase (423-832). Alternatively, the 5'→3' exonuclease domain and the 3'→5' exonuclease

domain of Tma DNA polymerase (ca. codons 1-484) may be fused to the DNA polymerase (dNTP binding and primer/template binding domains) portions of Taq DNA polymerase (ca. codons 423-832). The donors and recipients need not be limited to Taq and Tma DNA polymerases. Tth DNA polymerase provides analogous domains as Taq DNA polymerase. In addition, the enhanced/preferred reverse transcriptase properties of Tth DNA polymerase can be further enhanced by the addition of a 3'→5' exonuclease domain as illustrated above.

While any of a variety of means may be used to generate chimeric DNA polymerase coding sequences (possessing novel properties), a preferred method 10 employs "overlap" PCR. In this method, the intended junction sequence is designed into the PCR primers (at their 5'-ends). Following the initial amplification of the individual domains, the various products are diluted (ca. 100 to 1000-fold) and combined, denatured, annealed, extended, and then the final forward and reverse primers are added for an otherwise standard PCR.

15 Thus, the sequence that codes for the 3'→5' exonuclease activity of Tma DNA polymerase can be removed from Tma DNA polymerase or added to other polymerases that lack this activity by recombinant DNA methodology. One can even replace, in a non-thermostable DNA polymerase, the 3'→5' exonuclease activity domain with the thermostable 3'→5' exonuclease domain of Tma polymerase. Likewise, the 3'→5' 20 exonuclease activity domain of a non-thermostable DNA polymerase can be used to replace the 3'→5' exonuclease domain of Tma polymerase (or any other thermostable polymerase) to create a useful polymerase of the invention. Those of skill in the art recognize that the above chimeric polymerases are most easily constructed by recombinant DNA techniques. Similar chimeric polymerases can be constructed by 25 moving the 5'→3' exonuclease domain of one DNA polymerase to another.

30 Whether one desires to produce an enzyme identical to native Tma DNA polymerase or a derivative or homologue of that enzyme, the production of a recombinant form of Tma polymerase typically involves the construction of an expression vector, the transformation of a host cell with the vector, and culture of the transformed host cell under conditions such that expression will occur.

35 To construct the expression vector, a DNA is obtained that encodes the mature (used here to include all chimeras or muteins) enzyme or a fusion of the Tma polymerase to an additional sequence that does not destroy activity or to an additional sequence cleavable under controlled conditions (such as treatment with peptidase) to give an active protein. The coding sequence is then placed in operable linkage with suitable control sequences in an expression vector. The vector can be designed to replicate autonomously in the host cell or to integrate into the chromosomal DNA of the host cell. The vector is used to transform a suitable host, and the transformed host is cultured under conditions suitable for expression of recombinant Tma polymerase. The

Tma polymerase is isolated from the medium or from the cells, although recovery and purification of the protein may not be necessary in some instances.

Each of the foregoing steps can be done in a variety of ways. For example, the desired coding sequence may be obtained from genomic fragments and used directly in appropriate hosts. The construction for expression vectors operable in a variety of hosts is made using appropriate replicons and control sequences, as set forth generally below. Construction of suitable vectors containing the desired coding and control sequences employs standard ligation and restriction techniques that are well understood in the art. Isolated plasmids, DNA sequences, or synthesized oligonucleotides are cleaved, modified, and religated in the form desired. Suitable restriction sites can, if not normally available, be added to the ends of the coding sequence so as to facilitate construction of an expression vector, as exemplified below.

Site-specific DNA cleavage is performed by treating with suitable restriction enzyme (or enzymes) under conditions that are generally understood in the art and specified by the manufacturers of commercially available restriction enzymes. See, e.g., New England Biolabs, Product Catalog. In general, about 1  $\mu$ g of plasmid or other DNA is cleaved by one unit of enzyme in about 20  $\mu$ l of buffer solution; in the examples below, an excess of restriction enzyme is generally used to ensure complete digestion of the DNA. Incubation times of about one to two hours at about 37°C are typical, although variations can be tolerated. After each incubation, protein is removed by extraction with phenol and chloroform; this extraction can be followed by ether extraction and recovery of the DNA from aqueous fractions by precipitation with ethanol. If desired, size separation of the cleaved fragments may be performed by polyacrylamide gel or agarose gel electrophoresis using standard techniques. See, e.g., Methods in Enzymology, 1980, 65:499-560.

Restriction-cleaved fragments with single-strand "overhanging" termini can be made blunt-ended (double-strand ends) by treating with the large fragment of E. coli DNA polymerase I (Klenow) in the presence of the four deoxynucleoside triphosphates (dNTPs) using incubation times of about 15 to 25 minutes at 20°C to 25°C in 50 mM Tris pH 7.6, 50 mM NaCl, 10 mM MgCl<sub>2</sub>, 10 mM DTT, and 5 to 10  $\mu$ M dNTPs. The Klenow fragment fills in at 5' protruding ends, but chews back protruding 3' single strands, even though the four dNTPs are present. If desired, selective repair can be performed by supplying only one of the, or selected, dNTPs within the limitations dictated by the nature of the protruding ends. After treatment with Klenow, the mixture is extracted with phenol/chloroform and ethanol precipitated. Similar results can be achieved using S1 nuclease, because treatment under appropriate conditions with S1 nuclease results in hydrolysis of any single-stranded portion of a nucleic acid.

Synthetic oligonucleotides can be prepared using the triester method of Matteucci *et al.*, 1981, J. Am. Chem. Soc. 103:3185-3191, or automated synthesis

methods. Kinasing of single strands prior to annealing or for labeling is achieved using an excess, e.g., approximately 10 units, of polynucleotide kinase to 0.5  $\mu$ M substrate in the presence of 50 mM Tris, pH 7.6, 10 mM MgCl<sub>2</sub>, 5 mM dithiothreitol (DTT), and 1 to 2  $\mu$ M ATP. If kinasing is for labeling of probe, the ATP will contain high specific 5 activity  $\gamma$ -<sup>32</sup>P.

Ligation are performed in 15-30  $\mu$ l volumes under the following standard conditions and temperatures: 20 mM Tris-Cl, pH 7.5, 10 mM MgCl<sub>2</sub>, 10 mM DTT, 33  $\mu$ g/ml BSA, 10 mM-50 mM NaCl, and either 40  $\mu$ M ATP and 0.01-0.02 (Weiss) units T4 DNA ligase at 0°C (for ligation of fragments with complementary single-stranded 10 ends) or 1 mM ATP and 0.3-0.6 units T4 DNA ligase at 14°C (for "blunt end" ligation). Intermolecular ligations of fragments with complementary ends are usually performed at 33-100  $\mu$ g/ml total DNA concentrations (5 to 100 nM total ends concentration). Intermolecular blunt end ligations (usually employing a 20 to 30 fold molar excess of linkers, optionally) are performed at 1  $\mu$ M total ends concentration.

In vector construction, the vector fragment is commonly treated with bacterial or 15 calf intestinal alkaline phosphatase (BAP or CIAP) to remove the 5' phosphate and prevent religation and reconstruction of the vector. BAP and CIAP digestion conditions are well known in the art, and published protocols usually accompany the commercially available BAP and CIAP enzymes. To recover the nucleic acid 20 fragments, the preparation is extracted with phenol-chloroform and ethanol precipitated to remove AP and purify the DNA. Alternatively, religation can be prevented by restriction enzyme digestion of unwanted vector fragments before or after ligation with the desired vector.

For portions of vectors or coding sequences that require sequence 25 modifications, a variety of site-specific primer-directed mutagenesis methods are available. The polymerase chain reaction (PCR) can be used to perform site-specific mutagenesis. In another technique now standard in the art, a synthetic oligonucleotide encoding the desired mutation is used as a primer to direct synthesis of a complementary nucleic acid sequence of a single-stranded vector, such as pBS13+, that 30 serves as a template for construction of the extension product of the mutagenizing primer. The mutagenized DNA is transformed into a host bacterium, and cultures of the transformed bacteria are plated and identified. The identification of modified vectors may involve transfer of the DNA of selected transformants to a nitrocellulose filter or other membrane and the "lifts" hybridized with kinased synthetic primer at a 35 temperature that permits hybridization of an exact match to the modified sequence but prevents hybridization with the original strand. Transformants that contain DNA that hybridizes with the probe are then cultured and serve as a reservoir of the modified DNA.

In the constructions set forth below, correct ligations for plasmid construction are confirmed by first transforming *E. coli* strain DG101 or another suitable host with the ligation mixture. Successful transformants are selected by ampicillin, tetracycline or other antibiotic resistance or sensitivity or by using other markers, depending on the mode of plasmid construction, as is understood in the art. Plasmids from the transformants are then prepared according to the method of Clewell *et al.*, 1969, Proc. Natl. Acad. Sci. USA 62:1159, optionally following chloramphenicol amplification (Clewell, 1972, J. Bacteriol. 110:667). Another method for obtaining plasmid DNA is described as the "Base-Acid" extraction method at page 11 of the Bethesda Research Laboratories publication Focus, volume 5, number 2, and very pure plasmid DNA can be obtained by replacing steps 12 through 17 of the protocol with CsCl/ethidium bromide ultracentrifugation of the DNA. The isolated DNA is analyzed by restriction enzyme digestion and/or sequenced by the dideoxy method of Sanger *et al.*, 1977, Proc. Natl. Acad. Sci. USA 74:5463, as further described by Messing *et al.*, 1981, Nuc. Acids Res. 9:309, or by the method of Maxam *et al.*, 1980, Methods in Enzymology 65:499.

The control sequences, expression vectors, and transformation methods are dependent on the type of host cell used to express the gene. Generally, prokaryotic, yeast, insect, or mammalian cells are used as hosts. Prokaryotic hosts are in general the most efficient and convenient for the production of recombinant proteins and are therefore preferred for the expression of Tma polymerase.

The prokaryote most frequently used to express recombinant proteins is *E. coli*. For cloning and sequencing, and for expression of constructions under control of most bacterial promoters, *E. coli* K12 strain MM294, obtained from the *E. coli* Genetic Stock Center under GCSC #6135, can be used as the host. For expression vectors with the  $P_{L}N_{RBS}$  control sequence, *E. coli* K12 strain MC1000 lambda lysogen, N<sub>7</sub>N<sub>53</sub>Cl<sub>857</sub> SusP<sub>80</sub>, ATCC 39531, may be used. *E. coli* DG116, which was deposited with the ATCC (ATCC 53606) on April 7, 1987, and *E. coli* KB2, which was deposited with the ATCC (ATCC 53075) on March 29, 1985, are also useful host cells. For M13 phage recombinants, *E. coli* strains susceptible to phage infection, such as *E. coli* K12 strain DG98, are employed. The DG98 strain was deposited with the ATCC (ATCC 39768) on July 13, 1984.

However, microbial strains other than *E. coli* can also be used, such as bacilli, for example *Bacillus subtilis*, various species of *Pseudomonas*, and other bacterial strains, for recombinant expression of Tma DNA polymerase. In such prokaryotic systems, plasmid vectors that contain replication sites and control sequences derived from the host or a species compatible with the host are typically used.

For example, *E. coli* is typically transformed using derivatives of pBR322, described by Bolivar *et al.*, 1977, Gene 2:95. Plasmid pBR322 contains genes for

ampicillin and tetracycline resistance. These drug resistance markers can be either retained or destroyed in constructing the desired vector and so help to detect the presence of a desired recombinant. Commonly used prokaryotic control sequences, i.e., a promoter for transcription initiation, optionally with an operator, along with a ribosome binding site sequence, include the  $\beta$ -lactamase (penicillinase) and lactose (lac) promoter systems (Chang *et al.*, 1977, *Nature* 198:1056), the tryptophan (trp) promoter system (Goeddel *et al.*, 1980, *Nuc. Acids Res.* 8:4057), and the lambda-derived  $P_L$  promoter (Shimatake *et al.*, 1981, *Nature* 292:128) and N-gene ribosome binding site (NRBS). A portable control system cassette is set forth in United States Patent No. 4,711,845, issued December 8, 1987. This cassette comprises a  $P_L$  promoter operably linked to the NRBS in turn positioned upstream of a third DNA sequence having at least one restriction site that permits cleavage within six bp 3' of the NRBS sequence. Also useful is the phosphatase A (phoA) system described by Chang *et al.* in European Patent Publication No. 196,864, published October 8, 1986.

15 However, any available promoter system compatible with prokaryotes can be used to construct a Tma expression vector of the invention.

The nucleotide sequence of the Tma insert may negatively affect the efficiency of the upstream ribosomal binding site, resulting in low levels of translated polymerase. The translation of the Tma gene can be enhanced by the construction of "translationally coupled" derivatives of the expression vectors. An expression vector can be constructed with a secondary translation initiation signal and short coding sequence just upstream of the Tma gene coding sequence such that the stop codon for the short coding sequence is "coupled" with the ATG start codon for the Tma gene coding sequence. A secondary translation initiation signal that efficiently initiates translation can be inserted upstream of the Tma gene start codon. Translation of the short coding sequence brings the ribosome into close proximity with the Tma gene translation initiation site, thereby enhancing translation of the Tma gene. For example, one expression system can utilize the translation initiation signal and first ten codons of the T7 bacteriophage major capsid protein (gene 10) fused in-frame to the last six codons of TrpE. The TGA (stop) codon for TrpE is "coupled" with the ATG (start) codon for the Tma gene coding sequence, forming the sequence TGATG. A one base frame-shift is required between translation of the short coding sequence and translation of the Tma coding sequence. These derivative expression vectors can be constructed by recombinant DNA methods.

35 The redundancy of the genetic code can also be related to a low translation efficiency. Typically, when multiple codons coding the same amino acid occur, one of the possible codons is preferentially used in an organism. Frequently, an organism accumulates the tRNA species corresponding to the preferred codons at a higher level than those corresponding to rarely used codons. If the pattern of codon usage differs

between Thermotoga maritima and the host cell, the tRNA species necessary for translation of the Tma polymerase gene may be in low abundance. In the Tma coding sequence, arginine is most frequently coded for by the "AGA" codon, whereas this codon is used at low frequency in E. coli genes, and the corresponding tRNA is present in low concentration in E. coli host cells. Consequently, the low concentration in the E. coli host cell of "Arg U" tRNA for the "AGA" codon may limit the translation efficiency of the Tma polymerase gene RNA in E. coli host cells. The efficiency of translation of the Tma coding sequence within an E. coli host cell may be improved by increasing the concentration of this Arg tRNA species by expressing multiple copies of this tRNA gene in the host cell.

In addition to bacteria, eucaryotic microbes, such as yeast, can also be used as recombinant host cells. Laboratory strains of Saccharomyces cerevisiae, Baker's yeast, are most often used, although a number of other strains are commonly available. While vectors employing the two micron origin of replication are common (Broach, 1983, Meth. Enz. 101:307), other plasmid vectors suitable for yeast expression are known (see, for example, Sunchcomb *et al.*, 1979, Nature 282:39; Tschempe *et al.*, 1980, Gene 10:157; and Clarke *et al.*, 1983, Meth. Enz. 101:300). Control sequences for yeast vectors include promoters for the synthesis of glycolytic enzymes (Hess *et al.*, 1968, J. Adv. Enzyme Reg. 7:149; Holland *et al.*, 1978, Biotechnology 17:4900; and Holland *et al.*, 1981, J. Biol. Chem. 256:1385). Additional promoters known in the art include the promoter for 3-phosphoglycerate kinase (Hitzeman *et al.*, 1980, J. Biol. Chem. 255:2073) and those for other glycolytic enzymes, such as glyceraldehyde 3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. Other promoters that have the additional advantage of transcription controlled by growth conditions are the promoter regions for alcohol dehydrogenase 2, isocytchrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and enzymes responsible for maltose and galactose utilization (Holland, *supra*).

Terminator sequences may also be used to enhance expression when placed at the 3' end of the coding sequence. Such terminators are found in the 3' untranslated region following the coding sequences in yeast-derived genes. Any vector containing a yeast-compatible promoter, origin of replication, and other control sequences is suitable for use in constructing yeast Tma expression vectors.

The Tma gene can also be expressed in eucaryotic host cell cultures derived from multicellular organisms. See, for example, Tissue Culture, Academic Press, Cruz and Patterson, editors (1973). Useful host cell lines include COS-7, COS-A2, CV-1, murine cells such as murine myelomas N51 and VERO, HeLa cells, and Chinese hamster ovary (CHO) cells. Expression vectors for such cells ordinarily include

promoters and control sequences compatible with mammalian cells such as, for example, the commonly used early and late promoters from Simian Virus 40 (SV 40) (Fiers *et al.*, 1978, *Nature* 273:113), or other viral promoters such as those derived from polyoma, adenovirus 2, bovine papilloma virus (BPV), or avian sarcoma viruses, or immunoglobulin promoters and heat shock promoters. A system for expressing DNA in mammalian systems using a BPV vector system is disclosed in U.S. Patent No. 4,419,446. A modification of this system is described in U.S. Patent No. 4,601,978. General aspects of mammalian cell host system transformations have been described by Axel, U.S. Patent No. 4,399,216. "Enhancer" regions are also important in optimizing expression; these are, generally, sequences found upstream of the promoter region. Origins of replication may be obtained, if needed, from viral sources. However, integration into the chromosome is a common mechanism for DNA replication in eucaryotes.

Plant cells can also be used as hosts, and control sequences compatible with plant cells, such as the nopaline synthase promoter and polyadenylation signal sequences (Depicker *et al.*, 1982, *J. Mol. Appl. Gen.* 1:561) are available. Expression systems employing insect cells utilizing the control systems provided by baculovirus vectors have also been described (Miller *et al.*, 1986, *Genetic Engineering* (Setlow *et al.*, eds., Plenum Publishing) 8:277-297). Insect cell-based expression can be accomplished in *Spodoptera frugiperda*. These systems can also be used to produce recombinant *Tma* polymerase.

Depending on the host cell used, transformation is done using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described by Cohen, 1972, *Proc. Natl. Acad. Sci. USA* 69:2110 is used for prokaryotes or other cells that contain substantial cell wall barriers. Infection with *Agrobacterium tumefaciens* (Shaw *et al.*, 1983, *Gene* 23:315) is used for certain plant cells. For mammalian cells, the calcium phosphate precipitation method of Graham and van der Eb, 1978, *Virology* 52:546 is preferred. Transformations into yeast are carried out according to the method of Van Solingen *et al.*, 1977, *J. Bact.* 130:946 and Hsiao *et al.*, 1979, *Proc. Natl. Acad. Sci. USA* 76:3829.

Once the *Tma* DNA polymerase has been expressed in a recombinant host cell, purification of the protein may be desired. Although a variety of purification procedures can be used to purify the recombinant thermostable polymerase of the invention, fewer steps may be necessary to yield an enzyme preparation of equal purity. Because *E. coli* host proteins are heat-sensitive, the recombinant thermostable *Tma* DNA polymerase can be substantially enriched by heat inactivating the crude lysate. This step is done in the presence of a sufficient amount of salt (typically 0.3 M ammonium sulfate) to ensure dissociation of the *Tma* DNA polymerase from the host DNA and to reduce ionic interactions of *Tma* DNA polymerase with other cell lysate

proteins. In addition, the presence of 0.3 M ammonium sulfate promotes hydrophobic interaction with a phenyl sepharose column. Hydrophobic interaction chromatography is a separation technique in which substances are separated on the basis of differing strengths of hydrophobic interaction with an uncharged bed material containing hydrophobic groups. Typically, the column is first equilibrated under conditions favorable to hydrophobic binding, such as high ionic strength. A descending salt gradient may then be used to elute the sample.

According to the invention, an aqueous mixture (containing either native or recombinant Tma DNA polymerase) is loaded onto a column containing a relatively strong hydrophobic gel such as phenyl sepharose (manufactured by Pharmacia) or Phenyl TSK (manufactured by Toyo Soda). To promote hydrophobic interaction with a phenyl sepharose column, a solvent is used that contains, for example, greater than or equal to 0.3 M ammonium sulfate, with 0.3 M being preferred, or greater than or equal to 0.5 M NaCl. The column and the sample are adjusted to 0.3 M ammonium sulfate in 50 mM Tris (pH 7.5) and 1.0 mM EDTA ("TE") buffer that also contains 0.5 mM DTT, and the sample is applied to the column. The column is washed with the 0.3 M ammonium sulfate buffer. The enzyme may then be eluted with solvents that attenuate hydrophobic interactions, such as decreasing salt gradients, ethylene or propylene glycol, or urea. For native Tma DNA polymerase, a preferred embodiment involves washing the column with 2 M urea and 20% ethylene glycol in TE-DTT.

For long-term stability, Tma DNA polymerase enzyme can be stored in a buffer that contains one or more non-ionic polymeric detergents. Such detergents are generally those that have a molecular weight in the range of approximately 100 to 250,000 daltons, preferably about 4,000 to 200,000 daltons, and stabilize the enzyme at a pH of from about 3.5 to about 9.5, preferably from about 4 to 8.5. Examples of such detergents include those specified on pages 295-298 of McCutcheon's Emulsifiers & Detergents, North American edition (1983), published by the McCutcheon Division of MC Publishing Co., 175 Rock Road, Glen Rock, NJ (USA) and copending Serial No. 387,003, filed July 28, 1989, each of which is incorporated herein by reference.

Preferably, the detergents are selected from the group comprising ethoxylated fatty alcohol ethers and lauryl ethers, ethoxylated alkyl phenols, octylphenoxy polyethoxy ethanol compounds, modified oxyethylated and/or oxypropylated straight-chain alcohols, polyethylene glycol monooleate compounds, polysorbate compounds, and phenolic fatty alcohol ethers. More particularly preferred are Tween 20, a polyoxyethylated (20) sorbitan monolaurate from ICI Americas Inc., Wilmington, DE, and Iconol NP-40, an ethoxylated alkyl phenol (nonyl) from BASF Wyandotte Corp. Parsippany, NJ.

The thermostable enzyme of this invention may be used for any purpose in which such enzyme activity is necessary or desired. In a particularly preferred

embodiment, the enzyme catalyzes the nucleic acid amplification reaction known as PCR. This process for amplifying nucleic acid sequences is disclosed and claimed in U.S. Patent Nos. 4,683,202 and 4,865,188, each of which is incorporated herein by reference. The PCR nucleic acid amplification method involves amplifying at least one specific nucleic acid sequence contained in a nucleic acid or a mixture of nucleic acids and in the most common embodiment, produces double-stranded DNA.

For ease of discussion, the protocol set forth below assumes that the specific sequence to be amplified is contained in a double-stranded nucleic acid. However, the process is equally useful in amplifying single-stranded nucleic acid, such as mRNA, although in the preferred embodiment the ultimate product is still double-stranded DNA. In the amplification of a single-stranded nucleic acid, the first step involves the synthesis of a complementary strand (one of the two amplification primers can be used for this purpose), and the succeeding steps proceed as in the double-stranded amplification process described below.

This amplification process comprises the steps of:

(a) contacting each nucleic acid strand with four different nucleoside triphosphates and two oligonucleotide primers for each specific sequence being amplified, wherein each primer is selected to be substantially complementary to the different strands of the specific sequence, such that the extension product synthesized from one primer, when separated from its complement, can serve as a template for synthesis of the extension product of the other primer, said contacting being at a temperature that allows hybridization of each primer to a complementary nucleic acid strand;

(b) contacting each nucleic acid strand, at the same time as or after step (a), with a DNA polymerase from Thermotoga maritima that enables combination of the nucleoside triphosphates to form primer extension products complementary to each strand of the specific nucleic acid sequence;

(c) maintaining the mixture from step (b) at an effective temperature for an effective time to promote the activity of the enzyme and to synthesize, for each different sequence being amplified, an extension product of each primer that is complementary to each nucleic acid strand template, but not so high as to separate each extension product from the complementary strand template;

(d) heating the mixture from step (c) for an effective time and at an effective temperature to separate the primer extension products from the templates on which they were synthesized to produce single-stranded molecules but not so high as to denature irreversibly the enzyme;

(e) cooling the mixture from step (d) for an effective time and to an effective temperature to promote hybridization of a primer to each of the single-stranded molecules produced in step (d); and

(f) maintaining the mixture from step (e) at an effective temperature for an effective time to promote the activity of the enzyme and to synthesize, for each different sequence being amplified, an extension product of each primer that is complementary to each nucleic acid template produced in step (d) but not so high as to separate each extension product from the complementary strand template. The effective times and temperatures in steps (e) and (f) may coincide, so that steps (e) and (f) can be carried out simultaneously. Steps (d)-(f) are repeated until the desired level of amplification is obtained.

The amplification method is useful not only for producing large amounts of a specific nucleic acid sequence of known sequence but also for producing nucleic acid sequences that are known to exist but are not completely specified. One need know only a sufficient number of bases at both ends of the sequence in sufficient detail so that two oligonucleotide primers can be prepared that will hybridize to different strands of the desired sequence at relative positions along the sequence such that an extension product synthesized from one primer, when separated from the template (complement), can serve as a template for extension of the other primer into a nucleic acid sequence of defined length. The greater the knowledge about the bases at both ends of the sequence, the greater can be the specificity of the primers for the target nucleic acid sequence and the efficiency of the process.

In any case, an initial copy of the sequence to be amplified must be available, although the sequence need not be pure or a discrete molecule. In general, the amplification process involves a chain reaction for producing, in exponential quantities relative to the number of reaction steps involved, at least one specific nucleic acid sequence given that (a) the ends of the required sequence are known in sufficient detail that oligonucleotides can be synthesized that will hybridize to them and (b) that a small amount of the sequence is available to initiate the chain reaction. The product of the chain reaction will be a discrete nucleic acid duplex with termini corresponding to the 5' ends of the specific primers employed.

Any nucleic acid sequence, in purified or nonpurified form, can be utilized as the starting nucleic acid(s), provided it contains or is suspected to contain the specific nucleic acid sequence one desires to amplify. The nucleic acid to be amplified can be obtained from any source, for example, from plasmids such as pBR322, from cloned DNA or RNA, or from natural DNA or RNA from any source, including bacteria, yeast, viruses, organelles, and higher organisms such as plants and animals. DNA or RNA may be extracted from blood, tissue material such as chorionic villi, or amniotic cells by a variety of techniques. See, e.g., Maniatis *et al.*, *supra*, pp. 280-281. Thus, the process may employ, for example, DNA or RNA, including messenger RNA, which DNA or RNA may be single-stranded or double-stranded. In addition, a DNA-RNA hybrid that contains one strand of each may be utilized. A mixture of any of these

nucleic acids can also be employed as can nucleic acids produced from a previous amplification reaction (using the same or different primers). The specific nucleic acid sequence to be amplified can be only a fraction of a large molecule or can be present initially as a discrete molecule, so that the specific sequence constitutes the entire 5 nucleic acid.

The sequence to be amplified need not be present initially in a pure form; the sequence can be a minor fraction of a complex mixture, such as a portion of the  $\beta$ -globin gene contained in whole human DNA (as exemplified in Saiki *et al.*, 1985, *Science* 230:1530-1534) or a portion of a nucleic acid sequence due to a particular 10 microorganism, which organism might constitute only a very minor fraction of a particular biological sample. The cells can be directly used in the amplification process after suspension in hypotonic buffer and heat treatment at about 90 C-100 C until cell lysis and dispersion of intracellular components occur (generally 1 to 15 minutes).

After the heating step, the amplification reagents may be added directly to the lysed 15 cells. The starting nucleic acid sequence can contain more than one desired specific nucleic acid sequence. The amplification process is useful not only for producing large amounts of one specific nucleic acid sequence but also for amplifying simultaneously more than one different specific nucleic acid sequence located on the same or different nucleic acid molecules.

20 Primers play a key role in the PCR process. The word "primer" as used in describing the amplification process can refer to more than one primer, particularly in the case where there is some ambiguity in the information regarding the terminal sequence(s) of the fragment to be amplified or where one employs the degenerate primer process of the invention. For instance, in the case where a nucleic acid 25 sequence is inferred from protein sequence information, a collection of primers containing sequences representing all possible codon variations based on degeneracy of the genetic code will be used for each strand. One primer from this collection will be sufficiently homologous with the end of the desired sequence to be amplified to be useful for amplification.

30 In addition, more than one specific nucleic acid sequence can be amplified from the first nucleic acid or mixture of nucleic acids, so long as the appropriate number of different oligonucleotide primers are utilized. For example, if two different specific nucleic acid sequences are to be produced, four primers are utilized. Two of the primers are specific for one of the specific nucleic acid sequences, and the other two 35 primers are specific for the second specific nucleic acid sequence. In this manner, each of the two different specific sequences can be produced exponentially by the present process.

A sequence within a given sequence can be amplified after a given number of amplification cycles to obtain greater specificity in the reaction by adding, after at least

one cycle of amplification, a set of primers that are complementary to internal sequences (i.e., sequences that are not on the ends) of the sequence to be amplified. Such primers can be added at any stage and will provide a shorter amplified fragment. Alternatively, a longer fragment can be prepared by using primers with non-complementary ends but 5 having some overlap with the primers previously utilized in the amplification.

Primers also play a key role when the amplification process is used for in vitro mutagenesis. The product of an amplification reaction where the primers employed are not exactly complementary to the original template will contain the sequence of the primer rather than the template, so introducing an in vitro mutation. In further cycles, 10 this mutation will be amplified with an undiminished efficiency because no further mispaired priming is required. The process of making an altered DNA sequence as described above could be repeated on the altered DNA using different primers to induce further sequence changes. In this way, a series of mutated sequences can gradually be produced wherein each new addition to the series differs from the last in a minor way, 15 but from the original DNA source sequence in an increasingly major way.

Because the primer can contain as part of its sequence a non-complementary sequence, provided that a sufficient amount of the primer contains a sequence that is complementary to the strand to be amplified, many other advantages can be realized. For example, a nucleotide sequence that is not complementary to the template sequence 20 (such as, e.g., a promoter, linker, coding sequence, etc.) may be attached at the 5' end of one or both of the primers and so appended to the product of the amplification process. After the extension primer is added, sufficient cycles are run to achieve the desired amount of new template containing the non-complementary nucleotide insert. This allows production of large quantities of the combined fragments in a relatively 25 short period of time (e.g., two hours or less) using a simple technique.

Oligonucleotide primers can be prepared using any suitable method, such as, for example, the phosphotriester and phosphodiester methods described above, or 30 automated embodiments thereof. In one such automated embodiment, diethylphosphoramidites are used as starting materials and can be synthesized as described by Beaucage *et al.*, 1981, Tetrahedron Letters 22:1859-1862. One method for synthesizing oligonucleotides on a modified solid support is described in U.S. Patent No. 4,458,066. One can also use a primer that has been isolated from a biological source (such as a restriction endonuclease digest).

No matter what primers are used, however, the reaction mixture must contain a template for PCR to occur, because the specific nucleic acid sequence is produced by using a nucleic acid containing that sequence as a template. The first step involves contacting each nucleic acid strand with four different nucleoside triphosphates and two oligonucleotide primers for each specific nucleic acid sequence being amplified or detected. If the nucleic acids to be amplified or detected are DNA, then the nucleoside triphosphates are usually dATP, dCTP, dGTP, and dTTP, although various nucleotide derivatives can also be used in the process. The concentration of nucleoside triphosphates can vary widely. Typically, the concentration is 50 to 200  $\mu$ M in each 5 dNTP in the buffer for amplification, and MgCl<sub>2</sub> is present in the buffer in an amount of 1 to 3 mM to activate the polymerase and increase the specificity of the reaction. 10 However, dNTP concentrations of 1 to 20  $\mu$ M may be preferred for some applications, such as DNA sequencing or generating radiolabeled probes at high specific activity.

The nucleic acid strands of the target nucleic acid serve as templates for the 15 synthesis of additional nucleic acid strands, which are extension products of the primers. This synthesis can be performed using any suitable method, but generally occurs in a buffered aqueous solution, preferably at a pH of 7 to 9, most preferably about 8. To facilitate synthesis, a molar excess of the two oligonucleotide primers is added to the buffer containing the template strands. As a practical matter, the amount of 20 primer added will generally be in molar excess over the amount of complementary strand (template) when the sequence to be amplified is contained in a mixture of complicated long-chain nucleic acid strands. A large molar excess is preferred to improve the efficiency of the process. Accordingly, primer:template ratios of at least 1000:1 or higher are generally employed for cloned DNA templates, and primer: 25 template ratios of about 10<sup>8</sup>:1 or higher are generally employed for amplification from complex genomic samples.

The mixture of template, primers, and nucleoside triphosphates is then treated according to whether the nucleic acids being amplified or detected are double- or single-stranded. If the nucleic acids are single-stranded, then no denaturation step need be 30 employed prior to the first extension cycle, and the reaction mixture is held at a temperature that promotes hybridization of the primer to its complementary target (template) sequence. Such temperature is generally from about 35°C to 65°C or more, preferably about 37°C to 60°C for an effective time, generally from a few seconds to five minutes, preferably from 30 seconds to one minute. A hybridization temperature 35 of 35°C to 70°C may be used for Tma DNA polymerase. Primers that are 15 nucleotides or longer in length are used to increase the specificity of primer hybridization. Shorter primers require lower hybridization temperatures.

The complement to the original single-stranded nucleic acids can be synthesized by adding Tma DNA polymerase in the presence of the appropriate buffer, dNTPs, and

one or more oligonucleotide primers. If an appropriate single primer is added, the primer extension product will be complementary to the single-stranded nucleic acid and will be hybridized with the nucleic acid strand in a duplex of strands of equal or unequal length (depending on where the primer hybridizes to the template), which may 5 then be separated into single strands as described above to produce two single, separated, complementary strands. A second primer would then be added so that subsequent cycles of primer extension would occur using both the original single-stranded nucleic acid and the extension product of the first primer as templates. Alternatively, two or more appropriate primers (one of which will prime synthesis 10 using the extension product of the other primer as a template) can be added to the single-stranded nucleic acid and the reaction carried out.

If the nucleic acid contains two strands, as in the case of amplification of a double-stranded target or second-cycle amplification of a single-stranded target, the strands of nucleic acid must be separated before the primers are hybridized. This strand 15 separation can be accomplished by any suitable denaturing method, including physical, chemical or enzymatic means. One preferred physical method of separating the strands of the nucleic acid involves heating the nucleic acid until complete (>99%) denaturation occurs. Typical heat denaturation involves temperatures ranging from about 80°C to 105°C for times generally ranging from about a few seconds to minutes, depending on 20 the composition and size of the nucleic acid. Preferably, the effective denaturing temperature is 90°C-100°C for a few seconds to 1 minute. Strand separation may also be induced by an enzyme from the class of enzymes known as helicases or the enzyme RecA, which has helicase activity and in the presence of riboATP is known to denature DNA. The reaction conditions suitable for separating the strands of nucleic acids with 25 helicases are described by Kuhn Hoffmann-Berling, 1978, CSH-Quantitative Biology 43:63, and techniques for using RecA are reviewed in Radding, 1982, Ann. Rev. Genetics 16:405-437. The denaturation produces two separated complementary strands of equal or unequal length.

If the double-stranded nucleic acid is denatured by heat, the reaction mixture is 30 allowed to cool to a temperature that promotes hybridization of each primer to the complementary target (template) sequence. This temperature is usually from about 35°C to 65°C or more, depending on reagents, preferably 37°C to 60°C. The hybridization temperature is maintained for an effective time, generally a few seconds to minutes, and preferably 10 seconds to 1 minute. In practical terms, the temperature is 35 simply lowered from about 95°C to as low as 37°C, and hybridization occurs at a temperature within this range.

Whether the nucleic acid is single- or double-stranded, the DNA polymerase from Thermotoga maritima can be added prior to or during the denaturation step or when the temperature is being reduced to or is in the range for promoting hybridization.

Although the thermostability of Tma polymerase allows one to add Tma polymerase to the reaction mixture at any time, one can substantially inhibit non-specific amplification by adding the polymerase to the reaction mixture at a point in time when the mixture will not be cooled below the stringent hybridization temperature. After hybridization, 5 the reaction mixture is then heated to or maintained at a temperature at which the activity of the enzyme is promoted or optimized, i.e., a temperature sufficient to increase the activity of the enzyme in facilitating synthesis of the primer extension products from the hybridized primer and template. The temperature must actually be sufficient to synthesize an extension product of each primer that is complementary to each nucleic 10 acid template, but must not be so high as to denature each extension product from its complementary template (i.e., the temperature is generally less than about 80°C to 90°C).

Depending on the nucleic acid(s) employed, the typical temperature effective for this synthesis reaction generally ranges from about 40°C to 80°C, preferably 50°C to 15 75°C. The temperature more preferably ranges from about 65°C to 75°C for Thermotoga maritima DNA polymerase. The period of time required for this synthesis may range from about 10 seconds to several minutes or more, depending mainly on the temperature, the length of the nucleic acid, the enzyme, and the complexity of the nucleic acid mixture. The extension time is usually about 30 seconds to a few minutes. 20 If the nucleic acid is longer, a longer time period is generally required for complementary strand synthesis.

The newly synthesized strand and the complement nucleic acid strand form a double-stranded molecule that is used in the succeeding steps of the amplification process. In the next step, the strands of the double-stranded molecule are separated by 25 heat denaturation at a temperature and for a time effective to denature the molecule, but not at a temperature and for a period so long that the thermostable enzyme is completely and irreversibly denatured or inactivated. After this denaturation of template, the temperature is decreased to a level that promotes hybridization of the primer to the complementary single-stranded molecule (template) produced from the previous step, 30 as described above.

After this hybridization step, or concurrently with the hybridization step, the temperature is adjusted to a temperature that is effective to promote the activity of the thermostable enzyme to enable synthesis of a primer extension product using as a template both the newly synthesized and the original strands. The temperature again 35 must not be so high as to separate (denature) the extension product from its template, as described above. Hybridization may occur during this step, so that the previous step of cooling after denaturation is not required. In such a case, using simultaneous steps, the preferred temperature range is 50°C to 70°C.

The heating and cooling steps involved in one cycle of strand separation, hybridization, and extension product synthesis can be repeated as many times as needed to produce the desired quantity of the specific nucleic acid sequence. The only limitation is the amount of the primers, thermostable enzyme, and nucleoside triphosphates present. Usually, from 15 to 30 cycles are completed. For diagnostic detection of amplified DNA, the number of cycles will depend on the nature of the sample and the initial target concentration in the sample. For example, fewer cycles will be required if the sample being amplified is pure. If the sample is a complex mixture of nucleic acids, more cycles will be required to amplify the signal sufficiently for detection. For general amplification and detection, the process is repeated about 15 times. When amplification is used to generate sequences to be detected with labeled sequence-specific probes and when human genomic DNA is the target of amplification, the process is repeated 15 to 30 times to amplify the sequence sufficiently so that a clearly detectable signal is produced, i.e., so that background noise does not interfere with detection.

No additional nucleosides, primers, or thermostable enzyme need be added after the initial addition, provided that no key reagent has been exhausted and that the enzyme has not become denatured or irreversibly inactivated, in which case additional polymerase or other reagent would have to be added for the reaction to continue. Addition of such materials at each step, however, will not adversely affect the reaction. After the appropriate number of cycles has been completed to produce the desired amount of the specific nucleic acid sequence, the reaction can be halted in the usual manner, e.g., by inactivating the enzyme by adding EDTA, phenol, SDS, or CHCl<sub>3</sub> or by separating the components of the reaction.

The amplification process can be conducted continuously. In one embodiment of an automated process, the reaction mixture can be temperature cycled such that the temperature is programmed to be controlled at a certain level for a certain time. One such instrument for this purpose is the automated machine for handling the amplification reaction developed and marketed by Perkin-Elmer Cetus Instruments.

Detailed instructions for carrying out PCR with the instrument are available upon purchase of the instrument.

Tma DNA polymerase is very useful in the diverse processes in which amplification of a nucleic acid sequence by the polymerase chain reaction is useful. The amplification method may be utilized to clone a particular nucleic acid sequence for insertion into a suitable expression vector, as described in U.S. Patent No. 4,800,159. The vector may be used to transform an appropriate host organism to produce the gene product of the sequence by standard methods of recombinant DNA technology. Such cloning may involve direct ligation into a vector using blunt-end ligation, or use of restriction enzymes to cleave at sites contained within the primers. Other processes

suitable for Tma polymerase include those described in U.S. Patent Nos. 4,683,195 and 4,683,202 and European Patent Publication Nos. 229,701; 237,362; and 258,017; these patents and publications are incorporated herein by reference. In addition, the present enzyme is useful in asymmetric PCR (see Gyllensten and Erlich, 1988, *Proc. Natl. Acad. Sci. USA* 85:7652-7656, incorporated herein by reference); inverse PCR (Ochman *et al.*, 1988, *Genetics* 120:621, incorporated herein by reference); and for DNA sequencing (see Innis *et al.*, 1988, *Proc. Natl. Acad. Sci. USA* 85:9436-9440, and McConlogue *et al.*, 1988, *Nuc. Acids Res.* 16(20):9869). Tma polymerase is also believed to have reverse transcriptase activity; see PCT Patent Publication No. 10 91/09944, published July 11, 1991, incorporated herein by reference.

The reverse transcriptase activity of Tma DNA polymerase permits this enzyme to be used in methods for transcribing and amplifying RNA. The improvement of such methods resides in the use of a single enzyme, whereas previous methods have required more than one enzyme.

15 The improved methods comprise the steps of: (a) combining an RNA template with a suitable primer under conditions whereby the primer will anneal to the corresponding RNA template; and (b) reverse transcribing the RNA template by incubating the annealed primer-RNA template mixture with Tma DNA polymerase under conditions sufficient for the DNA polymerase to catalyze the polymerization of 20 deoxyribonucleoside triphosphates to form a DNA sequence complementary to the sequence of the RNA template.

In another aspect of the above method, the primer that anneals to the RNA template may also be suitable for amplification by PCR. In PCR, a second primer that is complementary to the reverse transcribed cDNA strand provides a site for initiation of 25 synthesis of an extension product. As already discussed above, the Tma DNA polymerase is able to catalyze this extension reaction on a cDNA template.

In the amplification of an RNA molecule by Tma DNA polymerase, the first extension reaction is reverse transcription, in which a DNA strand is produced in the form of an RNA/cDNA hybrid molecule. The second extension reaction, using the 30 DNA strand as a template, produces a double-stranded DNA molecule. Thus, synthesis of a complementary DNA strand from an RNA template with Tma DNA polymerase provides the starting material for amplification by PCR.

When Tma DNA polymerase is used for nucleic acid transcription from an RNA template, the use of buffers that contain  $Mn^{2+}$  provide improved stimulation of Tma 35 reverse transcriptase activity compared to previously used,  $Mg^{2+}$  containing reverse transcription buffers. Consequently, increased cDNA yields also result from these methods.

As stated above, the product of RNA transcription by Tma DNA polymerase is an RNA/cDNA hybrid molecule. The RNA is then removed by heat denaturation or

any number of other known methods including alkali, heat, or enzyme treatment. The remaining cDNA strand then serves as a template for polymerization of a self-complementary strand, thereby providing a double-stranded cDNA molecule suitable for amplification or other manipulation. The second strand synthesis requires a

5 sequence specific primer and Tma DNA polymerase.

Following the synthesis of the second cDNA strand, the resultant double-stranded cDNA molecule can serve a number of purposes, including DNA sequencing, amplification by PCR, or detection of a specific nucleic acid sequence. Specific primers useful for amplification of a segment of the cDNA can be added subsequent to 10 the reverse transcription. Also, one can use a first set of primers to synthesize a specific cDNA molecule and a second nested set of primers to amplify a desired cDNA segment. All of these reactions are catalyzed by Tma DNA polymerase.

Tma DNA polymerase can also be used to simplify and improve methods for 15 detection of RNA target molecules in a sample. In these methods, Tma DNA polymerase catalyzes: (a) reverse transcription; (b) second strand cDNA synthesis; and, if desired (c) amplification by PCR. In addition to the improvement of only requiring a single enzyme, the use of Tma DNA polymerase in the described methods eliminates the previous requirement of two sets of incubation conditions that were necessary due to the use of different enzymes for each procedural step. The use of Tma DNA 20 polymerase provides RNA transcription and amplification of the resulting complementary DNA with enhanced specificity and with fewer steps than previous RNA cloning and diagnostic methods. These methods are adaptable for use in kits for laboratory or clinical analysis.

The RNA that is transcribed and amplified in the above methods can be derived 25 from a number of sources. The RNA template can be contained within a nucleic acid preparation from any organism, such as a viral or bacterial nucleic acid preparation. The preparation can contain cell debris and other components, purified total RNA, or purified mRNA. The RNA template can also be a population of heterogeneous RNA molecules in a sample. Furthermore, the target RNA can be contained in a biological 30 sample, and the sample can be a heterogeneous sample in which RNA is but a small portion. Examples of such biological samples include blood samples and biopsied tissue samples.

Although the primers used in the reverse transcription step of the above 35 methods are generally completely complementary to the RNA template, the primers need not be completely complementary. As in PCR, not every nucleotide of the primer must anneal to the template for reverse transcription to occur. For example, a non-complementary nucleotide sequence can be present at the 5' end of the primer with the remainder of the primer sequence being complementary to the RNA. Alternatively, non-complementary bases can be interspersed into the primer, provided that the primer

sequence has sufficient complementarity with the RNA template for hybridization to occur and allow synthesis of a complementary DNA strand.

5 The following examples are offered by way of illustration only and are by no means intended to limit the scope of the claimed invention. In these examples, all percentages are by weight if for solids and by volume if for liquids, unless otherwise noted, and all temperatures are given in degrees Celsius.

#### EXAMPLE 1

##### Purification of *Thermotoga maritima* DNA Polymerase

10 This example describes the isolation of T<sub>mar</sub> DNA polymerase from Thermotoga maritima. The DNA polymerase was assayed at various points during purification according to the method described for Taq polymerase with one modification (1 mM MgCl<sub>2</sub>) in Lawyer *et al.*, 1989, J. Biol. Chem. 264(11):6427-6437, incorporated herein by reference.

15 Typically, this assay is performed in a total volume of 50  $\mu$ l of a reaction mixture composed of 25 mM TAPS-HCl, pH 9.5 (20°C); 50 mM KCl; 1 mM MgCl<sub>2</sub>; 1 mM  $\beta$ -mercaptoethanol; 200  $\mu$ M in each of dATP, dGTP, and TTP; 100  $\mu$ M  $\alpha$ -<sup>32</sup>P-dCTP (0.03 to 0.07  $\mu$ Ci/nMol); 12.5  $\mu$ g of activated salmon sperm DNA; and polymerase. The reaction is initiated by addition of polymerase in diluent (diluent is composed of 10 mM Tris-HCl, pH 8.0, 50 mM KCl, 0.1 mM EDTA, 1 mg/ml autoclaved gelatin, 0.5% NP40, 0.5% Tween 20, and 1 mM  $\beta$ -mercaptoethanol), and the reaction is carried out at 75°C. For the calculations shown below, one assumes that the volume of the polymerase (and diluent) added is 5  $\mu$ l, and the total reaction volume is 50  $\mu$ l. After a 10 minute incubation, the reaction is stopped by adding 10  $\mu$ l of 60 mM EDTA. The reaction mixture is centrifuged, and 50  $\mu$ l of reaction mixture is transferred to 1.0 ml of 50  $\mu$ g/ml carrier DNA in 2 mM EDTA (at 0°C). An equal volume (1 ml) of 20% TCA, 2% sodium pyrophosphate is added and mixed. The mixture is incubated at 0°C for 15 to 20 minutes and then filtered through Whatman GF/C filters and extensively washed (6 x 5 ml) with a cold mixture containing 5% TCA and 1% pyrophosphate, followed by a cold 95% ethanol wash. The filters are then dried and the radioactivity counted. Background (minus enzyme) is usually 0.001% to 0.01% of input cpm. About 50 to 250 pmoles <sup>32</sup>P-dCTP standard is spotted for unit calculation. One unit is equal to 10 nmoles dNTP incorporated in 30 minutes at 75°C. Units are calculated as follows.

35

sample cpm - enzyme dil. cpm = pmole dCTP incorporated  
specific activity of dCTP  
(cpm/pmole)

$$\frac{\text{pmole incorporated} \times 3 \times \text{dilution factor} \times 4}{4.167 \times 10} = \text{units/ml}$$

The 4.167 factor results from counting only 5/6 (50  $\mu$ l) of the reaction volume after the stop solution is added (60  $\mu$ l).

5 All operations were carried out at 0°C to 4°C unless otherwise stated. All glassware was baked prior to use, and solutions used in the purification were autoclaved, if possible, prior to use.

10 About 50 g of frozen *Thermotoga maritima* strain MSB8 cells (provided by Prof. Dr. K. O. Stetter, Regensburg, Germany) were thawed in 25 ml of 3 x TE-DTT buffer (150 mM Tris-Cl, pH 7.5, 3 mM EDTA, and 3 mM dithiothreitol) containing 2.4 mM PMSF (from 144 mM stock in DMF) and homogenized at low speed with a magnetic stirrer. The thawed cells were lysed in an Aminco french pressure cell (8 - 20,000 psi). The lysate was diluted with additional 1 x TE-DTT buffer containing fresh 2.4 mM PMSF to final 5.5x cell wet weight and sonicated to reduce viscosity (40 15 to 100% output, 9 min., 50% duty cycle).

20 The resulting fraction, fraction I (275 ml) contained 5.31 g of protein and  $15.5 \times 10^4$  units of activity. Ammonium sulfate was added to 0.2 M (7.25 g) and the lysate stirred for 15 minutes on ice. Ammonium sulfate prevents the *Tma* DNA polymerase from binding to DNA in the crude lysate and reduces ionic interactions of 25 the DNA polymerase with other cell lysate proteins.

Empirical testing showed that 0.2% Polymin P (polyethyleneimine, PEI) precipitates  $\geq 92\%$  of the total nucleic acid. Polymin P (pH 7.5) was added slowly to 0.2% (5.49 ml of 10% PEI) and the slurry stirred 30 minutes on ice, then centrifuged at 30,000 x g at 4°C for 30 minutes. The supernatant was designated fraction II (246 ml) 25 and contained 3.05 g of protein and  $12.5 \times 10^4$  units of activity.

30 Fraction II was adjusted to 0.3 M ammonium sulfate by addition of 3.24 g solid ammonium sulfate to ensure complete binding of the DNA polymerase to phenyl sepharose. Fraction II was then loaded onto a 2.2 x 6.6 cm (25 ml) phenyl sepharose CL-4B (lot OM 08012, purchased from Pharmacia - LKB) column (equilibrated in TE containing 0.3 M ammonium sulfate and 0.5 mM DTT) at 38 ml/hr (10 ml/cm<sup>2</sup>/hr). All resins were equilibrated and recycled according to the manufacturer's recommendations.

35 The column was washed with 150 ml of the same buffer (A<sub>280</sub> to baseline), then with 90 ml TE containing 0.5 mM DTT (no ammonium sulfate), followed by a wash with 95 ml of 20% ethylene glycol in TE containing 0.5 mM DTT and finally, eluted with 2 M urea in TE containing 20% ethylene glycol and 0.5 mM DTT. When the column fractions were assayed, a large proportion of the activity was found in the flow-through and wash fractions, indicating that the capacity of the column had been exceeded. Approximately 70% of the DNA polymerase which had bound to this first

phenyl sepharose column eluted at low salt (with the TE-DTT wash), and the balance of the bound material eluted with 2 M urea in 20% ethylene glycol in TE-DTT wash.

The flow-through activity from the first phenyl sepharose column was 5 designated PSII load (226 ml) and contained 1.76 g protein. Fraction PSII load was applied to a second phenyl sepharose column (of the same lot and dimensions), and the run was repeated the same way. Again, the capacity of the column was exceeded, and activity was found to elute with both the low salt and 2 M urea washes. Only 10% of the bound DNA polymerase eluted with the TE-DTT wash; the major portion (~90%) eluted with the 2 M urea in 20% ethylene glycol in TE-DTT wash.

10 The flow-through activity from the second phenyl sepharose column was combined with the TE-DTT eluates from the first and second phenyl sepharose columns and adjusted to 0.3 M ammonium sulfate. This fraction (PSIII load, 259.4 ml) contained 831 mg protein and was reapplied to a third phenyl sepharose column of 50 ml bed volume at 10 ml/cm<sup>2</sup>/hr. This time, all of the applied activity was retained by 15 the column and only eluted with the 2 M urea in 20% ethylene glycol in TE-DTT wash.

20 All three urea eluates were separately concentrated ~3 to 4 fold on Amicon YM30 membranes and dialyzed into heparin sepharose loading buffer shortly after elution to avoid prolonged exposure to urea (to avoid carbamylation). The dialyzed and concentrated urea eluates were assayed for protein concentration and were found to vary greatly in their specific activity. Because the urea eluate from the second phenyl sepharose column contained the majority of the activity at significantly higher specific 25 activity (~8 x 10<sup>4</sup> units of activity at ~1000 units/mg protein) than the other two eluates, it was processed separately from them.

30 The dialyzed and concentrated phenyl sepharose II urea eluate was applied to a 5 ml bed volume heparin sepharose CL 6B (purchased from Pharmacia - LKB) column that had been equilibrated with 0.08 M KCl, 50 mM Tris-Cl, pH 7.5, 0.1 mM EDTA, 0.2% Tween 20, and 0.5 mM DTT. This column and all subsequent columns were run at 1 bed volume per hr. All of the applied DNA polymerase activity was retained by the column. The column was washed with 17 ml of the same buffer (A<sub>280</sub> to baseline) and eluted with 60 ml of a linear 80 to 500 mM KCl gradient in the same buffer.

35 Fractions (0.53 ml) eluting between 0.21 and 0.315 M KCl were analyzed by SDS-PAGE. The peak fractions eluting between 0.225 and 0.275 M KCl were pooled separately. The flanking fractions were kept to be combined later with other fractions. The pool of peak fractions (affigel I load) was diluted with affigel-blue buffer without KCl to reduce its ionic strength to 0.15 M KCl.

The affigel I load fraction contained 3.4 mg of protein and was applied to a 4.3 ml affigel-blue (purchased from BioRad) column, which had been equilibrated in 25 mM Tris-Cl, pH 7.5, 0.1 mM EDTA, 0.2% Tween 20, 0.5 mM DTT, and 0.15 M KCl. All of the applied Tma DNA polymerase was retained. The column was washed

with 15 ml of the same buffer and eluted with a 66 ml linear 0.15 to 0.7 M KCl gradient in the same buffer.

Fractions (0.58 ml) eluting between 0.34 and 0.55 M KCl were analyzed by SDS-PAGE and appeared to be >90% pure. The polymerase peak fractions were no longer contaminated with site-specific endonuclease (indicated by absence of lower-molecular-weight specific DNA fragments after one or twenty-two hours incubation at 65°C with 2 units of Tma polymerase using 600 ng of plasmid pLSG 1 (ccc-DNA)). The polymerase peak fractions eluting between 0.3 and 0.55 M were pooled and concentrated ~20-fold on an Amicon YM 30 membrane. This fraction was then diafiltrated into 2.5 x storage buffer (50 mM Tris-Cl, pH 7.5, 250 mM KCl, 0.25 mM EDTA, 2.5 mM DTT, and 0.5% Tween 20 [Pierce, Surfact-Amps]) and stored at 4°C.

The urea eluates from the first and third phenyl sepharose columns were combined with the flanking fractions from the first heparin sepharose column. This pool (HSII load) contained ~200 mg protein and was diluted with heparin sepharose buffer without KCl to adjust its ionic strength to 80 mM KCl. HSII load was applied to a 16 ml bed volume heparin sepharose column (equilibrated in 80 mM KCl, 50 mM Tris-Cl, pH 7.5, 0.1 mM EDTA, 0.2% Tween 20, and 0.5 mM DTT). No detectable polymerase activity appeared in the flow-through fractions.

The column was washed with 80 ml of the same buffer and eluted with a 200 ml linear 80 to 750 mM KCl gradient in the same buffer. Fractions (2 ml) eluting between 0.225 and 0.335 M KCl were combined, concentrated ~5-fold on an Amicon YM 30 membrane, and dialyzed into hydroxyapatite-buffer. This fraction (HA load) contained 9.3 mg protein and was loaded onto a 4 ml bed volume hydroxyapatite (high resolution HPT, purchased from Calbiochem) column that had been equilibrated in 10 mM potassium phosphate buffer, pH 7.5, 0.5 mM DTT, 0.1 mM EDTA, and 0.2% Tween 20. All of the applied DNA polymerase activity was retained by the column.

The column was washed with 12 ml of the same buffer and eluted with a 60 ml linear 10 to 500 mM potassium phosphate (pH 7.5) gradient. Fractions (0.8 ml) eluting between 0.105 and 0.230 M potassium phosphate were analyzed by SDS-PAGE. Compared to the affigel column I load fraction (which by SDS-PAGE appeared to be ~10 to 20% pure), these fractions were ~5-fold less pure. The DNA polymerase peak fractions eluting between 0.105 and 0.255 M potassium phosphate were combined, concentrated ~3-fold on an Amicon YM 30 membrane, and diafiltrated into affigel-blue buffer.

The affigel II load fraction was applied to a 3 ml bed volume affigel-blue column that had been equilibrated in affigel-blue buffer. No detectable DNA polymerase activity appeared in the flow-through fractions. The column was washed with 9 ml of the same buffer and eluted with a 50 ml linear 0.2 to 0.7 M KCl gradient in the same buffer. Fractions (0.58 ml) eluting between 0.33 and 0.505 M KCl were

analyzed by SDS-PAGE. Because the earlier eluting fractions looked slightly cleaner by their silver staining pattern, two pools were made. Fractions eluting between 0.31 and 0.4 M KCl were combined into pool I; fractions eluting between 0.4 and 0.515 M KCl were combined into pool II. The two pools were each separately concentrated ~7-fold on an Amicon YM 30 membrane.

5 All three affigel-blue pools still contained high levels of contaminating, non-specific nucleases. Upon incubation at 70°C with 1.5 units of DNA polymerase, both a single-strand M13 DNA template and a multifragment restriction digest of a plasmid were degraded within a few hours. *In situ*-activity gels were run and showed that the 10 DNA polymerase fractions had not suffered proteolytic degradation.

The two pools from the second affigel-blue column were combined and dialyzed into a phosphocellulose column buffer. The dialyzed fraction (Pll I load) was loaded onto a 3 ml phosphocellulose column, which had been washed overnight with 25 mM Tris-Cl, pH 7.5, 50 mM KCl, 0.1 mM EDTA, 0.2% Tween 20, and 0.5 mM 15 DTT. This wash later proved to have been insufficient to equilibrate the pH of the phosphocellulose resin. Unfortunately, this was discovered after the sample had been loaded onto the column. All of the applied activity bound to the column.

The column was washed with 9 ml of loading buffer and eluted with a 45 ml linear 50 to 700 mM KCl gradient. DNA polymerase peak fractions (0.58 ml) eluting 20 between 0.46 and 0.575 M KCl were analyzed by SDS-PAGE.

Separation of contaminating proteins was observed throughout the peak: a ~45 kDa contaminating band elutes at 0.53 M KCl; an ~85 kDa contaminating band has an elution peak at 0.54 M KCl. Therefore, this column was repeated (loading at 25 somewhat higher ionic strength considering the elution profile of the polymerase). The peak fractions, eluting between 0.475 and 0.56 M KCl from the first phosphocellulose column were combined with the pool from the first affigel column. The combined fraction (Pll II load) now contained all of the purified polymerase (~7.5 x 10<sup>4</sup> units).

Fraction Pll II load was diluted with phosphocellulose buffer to adjust its ionic strength to 0.2 M KCl. Pll II load was loaded onto a 9 ml bed volume 30 phosphocellulose column, which, this time, had been equilibrated to the correct pH and ionic strength of 25 mM Tris-Cl, pH 7.5, 200 mM KCl, 0.1 mM EDTA, 0.2% Tween 20, and 0.5 mM DTT. The column was washed with 27 ml of the same buffer and was intended to be eluted with a 140 ml linear 0.2 to 0.8 M KCl gradient. However, instead of an upper limit buffer of 0.8 M KCl, the buffer had a concentration of 52 mM 35 KCl which resulted in a gradient decreasing in salt. The column was then reequilibrated with 32 ml of 0.2 M KCl-phosphocellulose buffer, and the 140 ml linear 0.2 to 0.8 M KCl gradient was reapplied.

The routine assays of flow-through, wash, and gradient fractions showed that, at this higher pH (pH 7.5), the DNA polymerase does not bind to the phosphocellulose

resin at 0.2 M KCl. The DNA polymerase activity containing fractions from the flow-through, wash, and decreasing salt-gradient-fractions were combined. The resulting pool was concentrated on an Amicon YM30 membrane. However, a mishap with the concentrator led to further losses of DNA polymerase activity. The recovered activity 5 was dialyzed into phosphocellulose buffer with 50 mM KCl and designated PlI III load.

This fraction was loaded onto a 5 ml bed volume phosphocellulose column that had been equilibrated with phosphocellulose buffer with 50 mM KCl. All of the applied activity was retained by the column. The column was washed with 15 ml of the same buffer and eluted with a 45 ml linear 50 to 500 mM KCl gradient in the same 10 buffer. Fractions (0.87 ml) eluting between 0.16 and 0.33 M KCl were analyzed by SDS-PAGE and *in situ* activity gels.

Based on the silver staining pattern, two pools were made. The peak fractions, eluting between 0.215 and 0.31 M KCl, were kept separate from the leading and trailing fractions, which were combined into a side-fractions pool. Both pools were 15 concentrated on centicon 30 membranes and diafiltered into 2.5x storage buffer (50 mM Tris-HCl, pH 7.5, 250 mM KCl, 0.25 mM EDTA, 2.5 mM DTT, and 0.5% Tween 20 [Pierce, Surfact-Amps]) and subsequently mixed with 1.5 volumes of 80% glycerol.

About  $3.1 \times 10^4$  units were recovered in the peak fraction; the side pool yields 20 an additional  $1 \times 10^3$  units of activity. The purified DNA polymerase was undegraded as evidenced by an unchanged migration pattern in an *in situ* activity gel. The molecular weight as determined by gel electrophoresis of the purified DNA polymerase is approximately 97 kDa. Tma DNA polymerase is recognized by epitope-specific 25 antibodies that correspond to Taq DNA polymerase amino acid residues number 569 through 587 (DGTP1) and 718 through 732 (DGTP3).

## EXAMPLE 2

### Isolation of DNA Encoding Tma

#### DNA Polymerase I Activity

Synthetic oligodeoxyribonucleotides DG164 through DG167 are four different 30 16-fold degenerate (each) 22mer pools designed as "forward" primers to one of the motifs in the template binding domains (3'-most 14 nucleotides) of thermostable DNA polymerases. This motif is the amino acid sequence Gly-Tyr-Val-Glu-Thr and corresponds identically to the T. aquaticus (Taq) DNA polymerase amino acids 718 through 722 and to the T. thermophilus (Tth) DNA polymerase amino acids 720 through 724. This motif is found in a DNA polymerase gene in all Thermus species. 35 The combined primer pool is 64-fold degenerate, and the primers encode a BglII recognition sequence at their 5'-ends.

Forward primers DG164 through DG167 are shown below:

DG164 SEQ ID NO: 2 5'CGAGATCTGGNTAYGTWGAAAC  
DG165 SEQ ID NO: 3 5'CGAGATCTGGNTAYGTWGAGAC  
DG166 SEQ ID NO: 4 5'CGAGATCTGGNTAYGTSGAAAC  
DG167 SEQ ID NO: 5 5'CGAGATCTGGNTAYGTSGAGAC

5 In these forward primers: A is Adenine; C is Cytidine; G is Guanidine; T is Thymine; Y is C+T (pYrimidine); S is G+C (Strong interaction; 3 H-bonds); W is A+T (Weak interaction; 2 H-bonds); and N is A+C+G+T (aNy).

10 Synthetic oligodeoxyribonucleotides DG160 through DG163 are four different 8-fold degenerate (each) 20mer pools designed as "reverse" primers to one of the motifs in the template binding domains (3'-most 14 nucleotides) of thermostable DNA polymerases. These primers are designed to complement the (+)-strand DNA sequence that encodes the motif Gln-Val-His-Asp-Glu and that corresponds identically to the Taq DNA polymerase amino acids 782 through 786 and to the Tth DNA polymerase amino acids 784 through 788. This motif is found in a DNA polymerase gene in all Thermus species. The combined primer pool is 32-fold degenerate, and the primers encode an EcoRI recognition sequence at their 5'-ends.

15

Reverse primers DG160 through 163 are shown below:

DG160 SEQ ID NO: 6 5'CGGAATTCRTCRTGWACCTG  
DG161 SEQ ID NO: 7 5'CGGAATTCRTCRTGWACTTG  
20 DG162 SEQ ID NO: 8 5'CGGAATTCRTCRTGSACCTG  
DG163 SEQ ID NO: 9 5'CGGAATTCRTCRTGSACTTG

In these reverse primers A, C, G, T, S, and W are as defined above, and R is G+A (puRine).

25 To amplify an ~230bp fragment of the Tma DNA polymerase gene, a PCR amplification tube was prepared without MgCl<sub>2</sub> that contained in 80 μl: (1) 5 ng denatured Tma genomic DNA; (2) 50 pmoles (total) of the combined forward primer set DG164-DG167; (3) 50 pmoles (total) of the combined reverse primer set DG160-DG163; (4) 2 units Taq DNA polymerase; (5) 50 μM each (final) dNTP; (6) 0.05% Laureth-12; and (7) standard PCR buffer except no magnesium chloride.

30 The sample was flash-frozen at -70°C and then stored at -20°C. The frozen sample was carefully layered with 20 μl of 10mM MgCl<sub>2</sub> (final concentration 2 mM), immediately overlayed with 50 μl of mineral oil, and cycled in a Perkin Elmer Cetus Thermal Cycler according to the following file: (1) step to 98°C - hold 50 seconds; (2) step to 50°C - hold 10 seconds; (3) ramp to 75°C over 4 minutes; and (4) step to 98°C.

35 The file was repeated for a total of 30 cycles. One-fifth (20 μl) of the amplification product was purified on a 3% Nusieve/1% Seakem agarose composite gel, and the approximately 230 bp fragment was eluted, concentrated, and digested with BglII and EcoRI.

Synthetic oligodeoxyribonucleotides DG154 and DG155 are two different 32-fold degenerate (each) 19mer pools designed as "forward" primers to one of the motifs in the primer:template binding domains (3'-most 11 nucleotides) of thermostable DNA polymerases. This motif is the tetrapeptide sequence Thr-Ala-Thr-Gly and corresponds identically to the Taq DNA polymerase amino acids 569 through 572 and to Tth DNA polymerase amino acids 571 through 574. This motif is found in a DNA polymerase gene in all Thermus species. The combined primer pool is 64-fold degenerate and the primers encode a BglII recognition sequence at their 5'-ends.

Forward primers DG154 and DG155 are presented below:

10 DG154 SEQ ID NO: 10 CGAGATCTACNGCNACWGG  
DG155 SEQ ID NO: 11 CGAGATCTACNGCNACSGG

In these forward primers, A, C, G, T, S, W, and N are as defined above.

15 To amplify an approximately ~667bp fragment of the Tma DNA polymerase gene, a PCR amplification tube was prepared without MgCl<sub>2</sub> that contained, in 80  $\mu$ l: (1) 5 ng denatured Tma genomic DNA; (2) 50 pmoles (total) of the combined forward primer set DG154-DG155; (3) 50 pmoles (total) of the combined reverse primer set DG160-DG163; (4) 2 Units of Taq DNA polymerase; (5) 50  $\mu$ M each (final) dNTP; (6) 0.05% Laureth 12; and (7) standard PCR buffer except no magnesium chloride.

20 The sample was flash-frozen at -70°C and then stored at -20°C. The frozen sample was carefully layered with 20  $\mu$ l of 10 mM MgCl<sub>2</sub> (final concentration 2 mM), immediately overlayed with 50  $\mu$ l of mineral oil, and cycled in a Perkin Elmer Cetus Thermal Cycler according to the following file: (1) step to 98°C - hold 50 seconds; (2) step to 55°C - hold 10 seconds; (3) ramp to 75°C over 4 minutes; (4) step to 98°C. The file was repeated for a total of 30 cycles.

25 One-fifth (20  $\mu$ l) of the amplification product was purified on a 1.5% agarose gel, and the approximately 670 bp fragment was eluted, concentrated, and digested with BglIII and EcoRI as above.

30 These amplification reactions yielded a 667 bp fragment and a 230 bp fragment, which was a subfragment of the 667 bp fragment. These fragments proved useful in obtaining the complete coding sequence for the Tma DNA polymerase I gene, as described in the following example.

### EXAMPLE 3

#### Cloning the *Thermotoga maritima* (Tma)

#### DNA Polymerase I Gene

35 This Example describes the strategy and methodology for cloning the Tma DNA polymerase I (Tma Pol I) gene of Thermotoga maritima.

The DNA sequences of the PCR products generated with primers DG164-167 and DG160-163 (230 bp) and DG154, 155 and DG160-163 (667 bp) contain an XbaI

restriction site recognition sequence, 5'CCCGGG. Oligonucleotides were designed to hybridize to sequences upstream and downstream of the XmaI site. DG224 is a 21mer, homologous to the PCR products 59-79 bp 3'-distal to the XmaI site. DG225 is a 22mer, homologous to the PCR products from the XmaI site to 21bp upstream (5') of the XmaI site. The sequence of DG224 and of DG225 is shown below (K is G or T).

5 DG224 SEQ ID NO: 12 5'ACAGCAGCKGATATAATAAG

DG225 SEQ ID NO: 13 5'GCCATGAGCTGTGGTATGTCTC

DG224 and DG225 were labelled by tailing with biotin-dUTP and terminal transferase in reactions designed to add approximately 8 biotin-dUMP residues to the 3'-end of 10 oligonucleotides. These labelled oligonucleotides were used as probes in Southern blot analyses of restriction digests of genomic Tma DNA. A preliminary restriction map was generated based on the Southern analysis results, and the DNA sequences of the PCR products that were generated as described in Example 2.

The preliminary map showed that the entire Tma DNA polymerase gene is 15 contained in two XmaI fragments. Most of the gene, including the 5'-end, resides on an approximately 2.6 kb XmaI fragment. The remainder of the gene (and the 3'-end) resides on an approximately 4.2 kb XmaI fragment. The two XmaI fragments containing the entire Tma DNA polymerase gene were cloned into plasmid pBS13+ (also called pBSM13+) as described below.

20 About 40 micrograms of Tma genomic DNA were digested to completion with XmaI. The XmaI digest was size-fractionated via electroelution. Slot blot analyses of a small portion of each fraction, using  $\gamma$ -32P-ATP-kinased DG224 and DG225 probes, identified the fractions containing the 4.2 kb 3'-fragment (hybridizing with DG224) and the 2.6 kb 5'-fragment (hybridizing with DG225). Fractions were concentrated via 25 ethanol precipitation and then ligated with XmaI-digested pBS13+ (Stratagene). Ampicillin-resistant transformants were selected on nitrocellulose filters and the filters probed with  $\gamma$ -32P-ATP-kinased DG224 or DG225 probe as appropriate. Plasmid DNA was isolated from colonies that hybridized with probe. Restriction analysis was performed to confirm that fragments were as expected and to determine orientation of 30 fragments relative to the pBS13+ vector.

DNA sequence analysis of the cloned fragments was performed using the 35 "universal" and "reverse" sequencing primers (which prime in the vector, outside the restriction site polylinker region). In addition, for 5'-clones, the primers used to determine the DNA sequence of the DG154-155/DG160-163 667 bp PCR clone were employed. Preliminary DNA sequence analysis confirmed that the desired DNA fragments containing the Tma DNA polymerase gene had been cloned.

From the preliminary DNA sequence, further sequencing primers were designed to obtain DNA sequence of more internal regions of the fragments. In addition, to facilitate DNA sequence analysis, several deletions of the two XmaI

fragments were made. For both orientations of the 2.6 kb 5'-fragment, EcoRI, SacI, and XbaI digests were each diluted and ligated under conditions that favored intramolecular ligation, thus deleting DNA between the vector EcoRI, SacI, and XbaI sites and the corresponding sites in the Tma XmaI fragment. Such internal deletions 5 allow ready DNA sequence analysis using the "universal" or "reverse" sequencing primers.

Similarly, a deletion of the 4.2 kb 3'-fragment was made, fusing the BamHI site of the vector with the BglII site approximately 650 bp from the Tma Pol I internal XmaI site in that clone (BamHI and BglII have identical GATC cohesive ends that 10 ligate readily with one another). This deletion allows for DNA sequence analysis of the 3'-end of the Tma Pol I gene.

Restriction site analysis reveals that both the 2.6 kb 5'-fragment and the 4.2 kb 3'-fragment lack NcoI, NdeI, and AseI restriction sites. Knowing the ATG start and coding sequence of the Tma Pol I gene, one can design oligonucleotides that will alter 15 the DNA sequence at the ATG start to include an NcoI, NdeI, or AseI restriction site via oligonucleotide site-directed mutagenesis. In addition, the mutagenic oligonucleotides can be designed such that a deletion of sequences between the lac promoter in the pBS13+ vector and the beginning of the Tma Pol I gene is made concurrent with the inclusion of an NdeI or AseI recognition sequence at the ATG start.

20 The deletion of sequences between the lac promoter in the vector and start of the Tma Pc gene would also eliminate the XmaI restriction site in the deleted region, thus making it convenient to assemble the entire coding sequence in an expression plasmid using conventional skill in the art (see, e.g., synthesis protocols for pDG174 - pDG181 in copending Serial No. 455,967, filed December 22, 1989, incorporated herein by 25 reference, and Example 5).

#### EXAMPLE 4

##### PCR With Tma DNA Polymerase

About 1.25 units of the Tma DNA polymerase purified in Example 1 is used to amplify rRNA sequences from Tth genomic DNA. The reaction volume is 50  $\mu$ l, and 30 the reaction mixture contains 50 pmol of primer DG73, 10<sup>5</sup> to 10<sup>6</sup> copies of the Tth genome (~2 x 10<sup>5</sup> copies of genome/ng DNA), 50 pmol of primer DG74, 200  $\mu$ M of each dNTP, 2 mM MgCl<sub>2</sub>, 10 mM Tris-HCl, pH 8.3, 50 mM KCl, and 100  $\mu$ g/ml gelatin (optionally, gelatin may be omitted).

35 The reaction is carried out on a Perkin-Elmer Cetus Instruments DNA Thermal Cycler. Twenty to thirty cycles of 96°C for 15 seconds; 50°C for 30 seconds, and 75°C for 30 seconds are carried out. At 20 cycles, the amplification product (160 bp in size) can be faintly seen on an ethidium bromide stained gel, and at 30 cycles, the product is readily visible (under UV light) on the ethidium bromide stained gel.

The PCR may yield fewer non-specific products if fewer units of Tma DNA polymerase are used (i.e., 0.31 units/50  $\mu$ l reaction). Furthermore, the addition of a non-ionic detergent, such as laureth-12, to the reaction mixture to a final concentration of about 0.5% to 1% can improve the yield of PCR product.

5        Primers DG73 and DG74 are shown below:

DG73      SEQ ID NO: 14    5' TACG<sub>1</sub>TCCCCGGCCTTGTAC  
DG74      SEQ ID NO: 15    5' AGGAGGTGATCCAACCGCA

#### EXAMPLE 5

##### Recombinant Expression

##### Vectors for Tma DNA Polymerase

##### A. Mutagenesis of the 5' and 3' Ends of the Tma Pol I Gene

The 5' end of the Tma gene in vector pBS:Tma7-1 (ATCC No. 68471, later renamed pTma01) was mutagenized with oligonucleotides DG240 and DG244 via oligonucleotide site-directed mutagenesis. Plasmid pBS:Tma7-1 consists of the 2.6 kb 15 5' XmaI fragment cloned into vector pBS13+. Resultant mutants from both mutageneses had deletions between the ATG of  $\beta$ -galactosidase in the pBS+ vector and the ATG of Tma Pol I so that the Tma coding sequence was positioned for expression utilizing the vector lac promoter, operator, and ribosome binding site (RBS). Both sets of mutants also had alterations in the second and sixth codons for Tma Pol I to be more compatible with the codon usage of E. coli without changing the amino acid sequence of the encoded protein. In addition, DG240 placed an NdeI restriction site at the ATG 20 start of the coding sequence (5'CATATG), and DG244 placed an NcoI restriction site at the ATG start of the coding sequence (5'CCATGG). DG240 mutant candidate colonies were screened with [ $\gamma$ <sup>32</sup>P]-labelled oligonucleotide DG241, and DG244 25 mutant candidate colonies were screened with [ $\gamma$ <sup>32</sup>P]-labelled oligonucleotide DG245. Plasmid DNA was isolated from colonies that hybridized with the appropriate probes, and mutations were confirmed via restriction analysis and DNA sequence analysis. The DG240 mutant was named pTma5'Nde#3 and later renamed pTma06. The DG244 mutant was named pTma5'Nco#9 and later renamed pTma07.

30        The 3'-end of the Tma Pol I gene was mutagenized in pBSTma3'11-1 Bam/BglI (ATCC No. 68472, later renamed pTma04) with mutagenic oligonucleotide DG238. Plasmid pBSTma3'11-1 Bam/BglI was constructed as described in Example 3 by cloning the 4.2 kb 3' XmaI fragment into pBS13+, digesting the resulting plasmid with BamHI and BglII, and circularizing by ligation the large fragment from the digestion. 35 DG238 inserts EcoRV and BamHI sites immediately downstream of the TGA stop codon. Mutant colony candidates were identified with [ $\gamma$ <sup>32</sup>P]-labelled oligonucleotide DG239. Plasmid DNA isolated from positive colonies was screened for appropriate

restriction digest patterns, and the DNA sequence was confirmed. One correct plasmid obtained was designated as pTma3'mut#1 and later renamed pTma05.

**B. Assembling the Full-Length Gene in a lac Promoter Vector**

For purposes of studying low level expression of Tma Pol I in E. coli and possible complementation of E. coli polymerase mutants by Tma Pol I (where high level expression might kill the cell, but where low level expression might rescue or complement), the Tma Pol I gene was assembled in the pBS13+ cloning vector. An ~300 bp XmaI to EcoRV fragment from pTma3'mut#1 was isolated and purified, following agarose gel electrophoresis and ethidium bromide staining, by excising an agarose gel slice containing the ~300 bp fragment and freezing in a Costar spinex filter unit. Upon thawing, the unit was spun in a microfuge, and the liquid containing the DNA fragment was collected. After ethanol precipitation, the fragment was ligated with each of the two 5'-mutated vectors, pTma5'Nde#3 and pTma5'Nco#9, which had each been digested with Asp718, repaired with Klenow and all 4 dNTPs (the reaction conditions are 56 mM Tris-Cl, pH 8.0, 56 mM NaCl, 6 mM MgCl<sub>2</sub>, 6 mM DTT, 1 mM dNTPs, and 11 units of Klenow at 37°C for 15 minutes; then inactivate at 75°C for 10 minutes), and then further digested with XmaI.

The ligation was carried out in two steps. To ligate the XmaI sticky ends, the conditions were 20 µg/ml total DNA, 20 mM Tris-Cl, pH 7.4, 50 mM NaCl, 10 mM MgCl<sub>2</sub>, 40 µM ATP, and 0.2 Weiss units T4 DNA ligase per 20 µl reaction at 0°C overnight. To ligate Asp718-digested, Klenow repaired blunt ends with EcoRV-digested blunt ends, the first ligations are diluted 4 to 5 fold and incubated at 15°C in the same ligation buffer, except 1 mM ATP and 10 Weiss units of T4 DNA ligase are used per 20 µl reaction. Ligations were transformed into DG101 host cells. Candidates were screened for appropriate restriction sites, and the DNA sequences around the cloning sites was confirmed. The desired plasmids were designated pTma08 (NdeI site at ATG) and pTma09 (NcoI site at ATG).

**C. Assembling the Full-Length Gene in P<sub>L</sub> Expression Vectors**

The following table describes P<sub>L</sub> promoter expression vectors used for assembling and expressing full-length Tma Pol I under the control of λ P<sub>L</sub> promoter.

Vector	Site at ATG	RBS*	Asp718+/-**	Oligonucleotide Duplexes	
				Cloned into pDG160 or pDG161	Amp/Tet***
pDG174	<u>Nde</u> I	T7	-	DG106/DG107	Amp
pDG178	<u>Nde</u> I	N	-	DG110/DG111	Amp
pDG182	<u>Nco</u> I	T7	+	FL42/FL43	Amp
pDG184	<u>Nco</u> I	N	+	FL44/FL45	Amp
pDG185	<u>Nco</u> I	N	+	FL44/FL45	Tet

- \*RBS - Phage T7 gene 10 or lambda gene N ribosome bind site.
- \*\* AsuII sites destroyed by digestion with Csp45I, repair with Klenow, and ligation of the repaired ends.
- \*\*\* Antibiotic resistance determinant ampicillin or tetracycline.

5 The five vectors in the table are derivatives of plasmid pDG160, if ampicillin resistant, or pDG161, if tetracycline resistant. Plasmids pDG160 and pDG161 and the scheme for constructing vectors similar to the pDG vectors shown in the table are described in Serial No. 455,967, filed December 22, 1989, incorporated herein by reference. The vectors confer ampicillin or tetracycline resistance and all contain the  $\delta$ -toxin positive  
10 retroregulator from Bacillus thuringiensis and the same point mutations in the RNA II gene that render the plasmids temperature sensitive for copy number.

The probes and oligonucleotides described in the Table are shown below.

DG240 SEQ ID NO: 16  
5'CCCATAAAAAGAAATAGTCTAGCCATATGTGTTCTGTGTGAAATTG  
15 DG241 SEQ ID NO: 17 5'AAACACATATGGCTAGAC  
DG244 SEQ ID NO: 18  
5'CCCATAAAAAGAAATAGTCTAGCCATGGTTGTTCTGTGTGAAATTG  
DG245 SEQ ID NO: 19 5'AAACAAACCATGGCTAGAC  
DG238 SEQ ID NO: 20  
20 5'GCAAAACATGGTCGTGATATCGGATCCGGAGGTGTTATCTGTGG  
DG239 SEQ ID NO: 21 5'CCGATATCACGACCATG  
DG106 SEQ ID NO: 22 5'CCGGAAGAAGGAGATACATATGAGCT  
DG107 SEQ ID NO: 23 5'CATATGTATATCTCCTTCTT  
DG110 SEQ ID NO: 24 5'CCGGAGGAGAAAACATATGAGCT  
25 DG111 SEQ ID NO: 25 5'CATATGTTTCTCCT  
FL42 SEQ ID NO: 26 5'CCGGAAGAAGGAGAAAATACCATGGGCCCGGTAC  
FL43 SEQ ID NO: 27 5'CGGGCCCATGGTATTTCTCCTTCTT  
FL44 SEQ ID NO: 28 5'CCGGAGGAGAAAATCCATGGGCCCGGTAC  
FL45 SEQ ID NO: 29 5'CGGGCCCATGGTATTTCTCCT  
30 A three-fragment ligation was used to assemble the Tma Pol I gene in the vectors. The vectors are digested with SmaI and either NdeI (pDG174, pDG178) or NcoI (pDG182, pDG184, pDG185). The 5' end of the Tma Pol I gene is from pTma5'Nde#3 digested with NdeI and XmaI or pTma5'Nco#9 digested with NcoI and XmaI. The 3' end of the gene is from pTma3'mut#1 digested with XmaI and EcoRV and the ~300 bp fragment purified as described above.  
35

The plasmid pDG182 shown in the Table and the scheme above were used to construct expression vector pTma13. The plasmid pDG184 and the scheme above were used to construct expression vectors pTma12-1 and pTma12-3. Plasmid pTma12-3 differs from pTma12-1 in that pTma12-3 is a dimer of pTma12-1 produced

during the same ligation/ transformation protocol. The plasmid pDG185 and the scheme shown above were used to construct expression vector pTma11.

Even though a vector may contain the entire polymerase coding sequence, a shortened form of the enzyme can be expressed either exclusively or in combination 5 with a full length polymerase. These shortened forms of Tma DNA polymerase result from translation initiation occurring at one of the methionine (ATG) codons in the coding sequence other than the 5'-ATG. The monomeric pTma12-1 plasmid produces, upon heat induction, predominantly a biologically active thermostable DNA polymerase lacking amino acids 1 through 139 of native Tma DNA polymerase. This 10 approximately 86 kDa protein is the result of translation initiation at the methionine codon at position 140 of the Tma coding sequence and is called MET140.

In shake flask studies under the appropriate conditions (heat induction at 34°C or 36°C, but not 38°C), the multimeric pTma12-3 expression vector yielded a significant level of "full length" Tma DNA polymerase (approximately 97 kDa by SDS- 15 PAGE) and a smaller amount of the shortened (approximately 86 kDa) form resulting from translation initiation at Met 140. Amino-acid sequencing of the full length Tma DNA polymerase indicated that the amino-terminal methionine was removed and the second-position alanine was present at the N-terminus.

Recombinant Tma DNA Polymerase was purified from E. coli strain DG116 20 containing plasmid pTma12-3. The seed flask for a 10 L fermentation contained tryptone (20 g/l), yeast extract (10 g/l), NaCl (10 g/l), ampicillin (100 mg/l), and thiamine (10 mg/l). The seed flask was inoculated with a colony from an agar plate (a frozen glycerol culture can be used). The seed flask was grown at 30°C to between 0.5 to 2.0 O.D. (A<sub>680</sub>). The volume of seed culture inoculated into the fermentor is 25 calculated such that the bacterial concentration is 0.5 mg dry weight/liter. The 12.5 liter growth medium contained 60 mM K<sub>2</sub>HPO<sub>4</sub>, 16 mM NaNH<sub>4</sub>HPO<sub>4</sub>, 10 mM citric acid, and 1 mM MgSO<sub>4</sub>. The following sterile components were added: 2 g/l glucose, 10 mg/l thiamine, 2.5 g/l casamino acids, 100 mg/l ampicillin, and 100 mg/l methicillin. Foaming was controlled by the addition of propylene glycol as necessary, as an 30 antifoaming agent. Airflow was maintained at 2 l/min.

The fermentor was inoculated as described above, and the culture was grown at 30°C for 4.5 hours to a cell density (A<sub>680</sub>) of 0.7. The growth temperature was shifted to 35°C to induce the synthesis of recombinant Tma DNA polymerase. The temperature shift increases the copy number of the pTma12-3 plasmid and simultaneously 35 derepresses the lambda P<sub>L</sub> promoter controlling transcription of the modified Tma DNA polymerase gene through inactivation of the temperature-sensitive cI repressor encoded by the defective prophage lysogen in the host. The cells were grown for 21 hours to an optical density of 4 (A<sub>680</sub>) and harvested by centrifugation. The resulting cell paste was stored at -70°C.

Recombinant Tma DNA polymerase is purified as in Example 6, below.

Briefly, cells are thawed in 1 volume of TE buffer (50 mM Tris-Cl, pH 7.5, and 1.0 mM EDTA with 1mM DTT), and protease inhibitors are added (PMSF to 2.4 mM, leupeptin to 1  $\mu$ g/ml, and TLCK to 0.2 mM). The cells are lysed in an Aminco french pressure cell at 20,000 psi and sonicated to reduce viscosity. The sonicate is diluted with TE buffer and protease inhibitors to 5.5 X wet weight cell mass (Fraction I), adjusted to 0.3 M ammonium sulfate, and brought rapidly to 75°C and maintained at 75°C for 15 min. The heat-treated supernatant is chilled rapidly to 0°C, and the E. coli cell membranes and denatured proteins are removed following centrifugation at 20,000 X G for 30 min. The supernatant containing Tma DNA polymerase (Fraction II) is saved. The level of Polymyxin P necessary to precipitate >95% of the nucleic acids is determined by trial precipitation (usually in the range of 0.6 to 1% w/v). The desired amount of Polymyxin P is added slowly with rapid stirring at 0°C for 30 min. and the suspension centrifuged at 20,000 X G for 30 min. to remove the precipitated nucleic acids. The supernatant (Fraction III) containing the Tma DNA polymerase is saved.

Fraction III is applied to a phenyl sepharose column that has been equilibrated in 50 mM Tris-Cl, pH 7.5, 0.3 M ammonium sulfate, 10 mM EDTA, and 1 mM DTT. The column is washed with 2 to 4 column volumes of the same buffer ( $A_{280}$  to baseline), and then 1 to 2 column volumes of TE buffer containing 100 mM KCl to remove most contaminating E. coli proteins. Tma DNA polymerase is then eluted from the column with buffer containing 50 mM Tris-Cl, pH 7.5, 2 M urea, 20% (w/v) ethylene glycol, 10 mM EDTA, and 1 mM DTT, and fractions containing DNA polymerase activity are pooled (Fraction IV).

Final purification of recombinant Tma DNA polymerase is achieved using heparin sepharose chromatography (as for native or MET284 recombinant DNA polymerase), anion exchange chromatography, or affigel blue chromatography. Recombinant Tma DNA polymerase may be diafiltered into 2.5X storage buffer, combined with 1.5 volumes of sterile 80% (w/v) glycerol, and stored at -20°C.

#### Example 6

##### Expression of a Truncated Tma Polymerase MET284

As noted above, expression plasmids containing the complete Tma gene coding sequence expressed either a full length polymerase resulting from translation initiation at the start codon or a shortened polymerase resulting from translation initiation occurring at the methionine codon at position 140. A third methionine codon that can act as a translation initiation site occurs at position 284 of the Tma gene coding sequence. Plasmids that express a DNA polymerase lacking amino acids 1 through 283 of native Tma DNA polymerase were constructed by introducing deleting corresponding regions of the Tma coding sequence.

Plasmid pTma12-1 was digested with BspHI (nucleotide position 848) and HindIII (nucleotide position 2629). A 1781 base pair fragment was isolated by agarose gel purification. To separate the agarose from the DNA, a gel slice containing the desired fragment was frozen at -20°C in a Costar spinex filter unit. After thawing at 5 room temperature, the unit was spun in a microfuge. The filtrate containing the DNA was concentrated in a Speed Vac concentrator, and the DNA was precipitated with ethanol.

The isolated fragment was cloned into plasmid pTma12-1 digested with NcoI and HindIII. Because NcoI digestion leaves the same cohesive end sequence as digestion with BspHI, the 1781 base pair fragment has the same cohesive ends as the full length fragment excised from plasmid pTma12-1 by digestion with NcoI and HindIII. The ligation of the isolated fragment with the digested plasmid results in a fragment switch and was used to create a plasmid designated pTma14.

Plasmid pTma15 was similarly constructed by cloning the same isolated 15 fragment into pTma13. As with pTma14, pTma15 drives expression of a polymerase that lacks amino acids 1 through 283 of native Tma DNA polymerase; translation initiates at the methionine codon at position 284 of the native coding sequence.

Both the pTma14 and pTma15 expression plasmids expressed at a high level a biologically active thermostable DNA polymerase of molecular weight of about 70 kDa; 20 plasmid pTma15 expressed polymerase at a higher level than did pTma14. Based on similarities with E. coli Pol I Klenow fragment, such as conservation of amino acid sequence motifs in all three domains that are critical for 3'-5' exonuclease activity, distance from the amino terminus to the first domain critical for exonuclease activity, and length of the expressed protein, the shortened form (MET284) of Tma polymerase 25 should possess 3'-5' exonuclease and proof-reading activity but lack 5'-3' exonuclease activity. However, initial SDS activity gel assays and solution assays for 3'-5' exonuclease activity suggested significant attenuation in the proof-reading activity of the polymerase expressed by E. coli host cells harboring plasmid pTma15.

MET284 Tma DNA Polymerase was purified from E. coli strain DG116 30 containing plasmid pTma15. The seed flask for a 10 L fermentation contained tryptone (20 g/l), yeast extract (10 g/l), NaCl (10 g/l), glucose (10 g/l), ampicillin (50 mg/l), and thiamine (10 mg/l). The seed flask was inoculated with a colony from an agar plate (a frozen glycerol culture can be used). The seed flask was grown at 30°C to between 0.5 to 2.0 O.D. (A<sub>680</sub>). The volume of seed culture inoculated into the fermentor is 35 calculated such that the bacterial concentration is 0.5 mg dry weight/liter. The 10 liter growth medium contained 25 mM KH<sub>2</sub>PO<sub>4</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 4 mM sodium citrate, 0.4 mM FeCl<sub>3</sub>, 0.04 mM ZnCl<sub>2</sub>, 0.03 mM CoCl<sub>2</sub>, 0.03 mM CuCl<sub>2</sub>, and 0.03 mM H<sub>3</sub>BO<sub>3</sub>. The following sterile components were added: 4 mM MgSO<sub>4</sub>, 20 g/l glucose, 20 mg/l thiamine, and 50 mg/l ampicillin. The pH was adjusted to 6.8 with NaOH and

controlled during the fermentation by added NH<sub>4</sub>OH. Glucose was continually added by coupling to NH<sub>4</sub>OH addition. Foaming was controlled by the addition of propylene glycol as necessary, as an antifoaming agent. Dissolved oxygen concentration was maintained at 40%.

5 The fermentor was inoculated as described above, and the culture was grown at 30°C to a cell density of 0.5 to 1.0 X 10<sup>10</sup> cells/ml (optical density [A<sub>680</sub>] of 15). The growth temperature was shifted to 38°C to induce the synthesis of MET284 Tma DNA polymerase. The temperature shift increases the copy number of the pTma15 plasmid and simultaneously derepresses the lambda P<sub>L</sub> promoter controlling transcription of the 10 modified Tma DNA polymerase gene through inactivation of the temperature-sensitive cI repressor encoded by the defective prophage lysogen in the host.

15 The cells were grown for 6 hours to an optical density of 37 (A<sub>680</sub>) and harvested by centrifugation. The cell mass (ca. 95 g/l) was resuspended in an equivalent volume of buffer containing 50 mM Tris-Cl, pH 7.6, 20 mM EDTA and 20% (w/v) glycerol. The suspension was slowly dripped into liquid nitrogen to freeze the suspension as "beads" or small pellets. The frozen cells were stored at -70°C.

20 To 200 g of frozen beads (containing 100 g wet weight cell) were added 100 ml of 1X TE (50 mM Tris-Cl, pH 7.5, 10 mM EDTA) and DTT to 0.3 mM, PMSF to 2.4 mM, leupeptin to 1 µg/ml and TLCK (a protease inhibitor) to 0.2 mM. The sample was thawed on ice and uniformly resuspended in a blender at low speed. The cell suspension was lysed in an Aminco french pressure cell at 20,000 psi. To reduce viscosity, the lysed cell sample was sonicated 4 times for 3 min. each at 50% duty cycle and 70% output. The sonicate was adjusted to 550 ml with 1X TE containing 1 mM DTT, 2.4 mM PMSF, 1 µg/ml leupeptin and 0.2 mM TLCK (Fraction I). After 25 addition of ammonium sulfate to 0.3 M, the crude lysate was rapidly brought to 75°C in a boiling water bath and transferred to a 75°C water bath for 15 min. to denature and inactivate E. coli host proteins. The heat-treated sample was chilled rapidly to 0°C and incubated on ice for 20 min. Precipitated proteins and cell membranes were removed by centrifugation at 20,000 X G for 30 min. at 5°C and the supernatant (Fraction II) 30 saved.

35 The heat-treated supernatant (Fraction II) was treated with polyethyleneimine (PEI) to remove most of the DNA and RNA. Polymin P (34.96 ml of 10% [w/v], pH 7.5) was slowly added to 437 ml of Fraction II at 0°C while stirring rapidly. After 30 min. at 0°C, the sample was centrifuged at 20,000 X G for 30 min. The supernatant (Fraction III) was applied at 80 ml/hr to a 100 ml phenylseparose column (3.2x12.5 cm) that had been equilibrated in 50 mM Tris-Cl, pH 7.5, 0.3 M ammonium sulfate, 10 mM EDTA, and 1 mM DTT. The column was washed with about 200 ml of the same buffer (A<sub>280</sub> to baseline) and then with 150 ml of 50 mM Tris-Cl, pH 7.5, 100 mM KCl, 10 mM EDTA and 1 mM DTT. The MET284 Tma DNA polymerase was then

eluted from the column with buffer containing 50 mM Tris-Cl, pH 7.5, 2 M urea, 20% (w/v) ethylene glycol, 10 mM EDTA, and 1 mM DTT, and fractions containing DNA polymerase activity were pooled (Fraction IV).

Fraction IV is adjusted to a conductivity equivalent to 50 mM KCl in 50 mM

5 Tris-Cl, pH 7.5, 1 mM EDTA, and 1 mM DTT. The sample was applied (at 9 ml/hr) to a 15 ml heparin-sepharose column that had been equilibrated in the same buffer. The column was washed with the same buffer at ca. 14 ml/hr (3.5 column volumes) and eluted with a 150 ml 0.05 to 0.5 M KCl gradient in the same buffer. The DNA polymerase activity eluted between 0.11-0.22 M KCl. Fractions containing the  
10 pTma15 encoded modified Tma DNA polymerase are pooled, concentrated, and diafiltered against 2.5X storage buffer (50 mM Tris-Cl, pH 8.0, 250 mM KCl, 0.25 mM EDTA, 2.5 mM DTT, and 0.5% Tween 20), subsequently mixed with 1.5 volumes of sterile 80% (w/v) glycerol, and stored at -20°C. Optionally, the heparin sepharose-eluted DNA polymerase or the phenyl sepharose-eluted DNA polymerase  
15 can be dialyzed or adjusted to a conductivity equivalent to 50 mM KCl in 50 mM Tris-Cl, pH 7.5, 1 mM DTT, 1 mM EDTA, and 0.2% Tween 20 and applied (1 mg protein/ml resin) to an affigel blue column that has been equilibrated in the same buffer. The column is washed with three to five column volumes of the same buffer and eluted with a 10 column volume KCl gradient (0.05 to 0.8 M) in the same buffer. Fractions  
20 containing DNA polymerase activity (eluting between 0.25 and 0.4 M KCl) are pooled, concentrated, diafiltered, and stored as above.

The relative thermostability of various DNA polymerases has been compared. At 97.5°C the half-life of native Tma DNA polymerase is more than twice the half-life of either native or recombinant Taq DNA (i.e., AmpliTaq<sup>®</sup>) DNA polymerase.

25 Surprisingly, the half-life at 97.5°C of MET284 Tma DNA polymerase is 2.5 to 3 times longer than the half-life of native Tma DNA polymerase.

PCR tubes containing 10 mM Tris-Cl, pH 8.3, and 1.5 mM MgCl<sub>2</sub> (for Taq or native Tma DNA polymerase) or 3 mM MgCl<sub>2</sub> (for MET284 Tma DNA polymerase), 50 mM KCl (for Taq, native Tma and MET284 Tma DNA polymerases) or no KCl (for  
30 MET284 Tma DNA polymerase), 0.5 μM each of primers PCR01 and PCR02, 1 ng of lambda template DNA, 200 μM of each dNTP except dCTP, and 4 units of each enzyme were incubated at 97.5°C in a large water bath for times ranging from 0 to 60 min. Samples were withdrawn with time, stored at 0°C, and 5 μl assayed at 75°C for 10 min. in a standard activity assay for residual activity.

35 Taq DNA polymerase had a half-life of about 10 min. at 97.5°C, while native Tma DNA polymerase had a half-life of about 21 to 22 min. at 97.5°C. Surprisingly, the MET284 form of Tma DNA polymerase had a significantly longer half-life (50 to 55 min.) than either Taq or native Tma DNA polymerase. The improved thermostability of MET284 Tma DNA polymerase will find applications in PCR,

particularly where G+C-rich targets are difficult to amplify because the strand-separation temperature required for complete denaturation of target and PCR product sequences leads to enzyme inactivation.

PCR tubes containing 50  $\mu$ l of 10 mM Tris-Cl, pH 8.3, 3 mM MgCl<sub>2</sub>, 200  $\mu$ M of each dNTP, 0.5 ng bacteriophage lambda DNA, 0.5  $\mu$ M of primer PCR01, 4 units of MET284 Tma DNA polymerase, and 0.5  $\mu$ M of primer PCR02 or PL10 were cycled 5 for 25 cycles using  $T_{den}$  of 96°C for 1 min. and  $T_{anneal-extend}$  of 60°C for 2 min. Lambda DNA template, deoxynucleotide stock solutions, and primers PCR01 and 10 PCR02 were part of the PECI GeneAmp® kit. Primer PL10 has the sequence: (SEQ ID NO. 45) 5'-GGCGTACCTTGTCTCACGGGCAAC-3' and is complementary to bacteriophage lambda nucleotides 8106-8130.

The primers PCR01 and PCR02 amplify a 500 bp product from lambda. The primer pair PCR01 and PL10 amplify a 1 kb product from lambda. After amplification with the respective primer sets, 5  $\mu$ l aliquots were subjected to agarose gel 15 electrophoresis and the specific intended product bands visualized with ethidium bromide staining. Abundant levels of product were generated with both primer sets, showing that MET284 Tma DNA polymerase successfully amplified the intended target sequence.

#### Example 7

##### Expression of Truncated Tma Polymerase

As noted above, host cells transformed with plasmids that contain the complete Tma DNA polymerase gene coding sequence express a shortened form (MET140) of Tma polymerase either exclusively or along with the full length polymerase. Mutations can be made to control which form of the polymerase is expressed. To enhance the 25 exclusive expression of the MET140 form of the polymerase, the coding region corresponding to amino acids through 139 were deleted from the expression vector. The protocol for constructing such a deletion is similar to the construction described in Example 6: a shortened gene fragment is excised and then reinserted into a vector from which a full length fragment has been excised. However, the shortened fragment can 30 be obtained as a PCR amplification product rather than purified from a restriction digest. This methodology allows a new upstream restriction site (or other sequences) to be incorporated where useful.

To delete the region up to the methionine codon at position 140, an SphI site was introduced into pTma12-1 and pTma13 using PCR. A forward primer (FL63) was 35 designed to introduce the SphI site just upstream of the methionine codon at position 140. The reverse primer (FL69) was chosen to include an XbaI at position 624. Plasmid pTma12-1 linearized with SmaI was used as the PCR template, yielding a 225 bp PCR product.

Before digestion, the PCR product was treated with 50 µg/ml of Proteinase K in PCR reaction mix plus 0.5% SDS and 5 mM EDTA. After incubating for 30 minutes at 37°C, the Proteinase K was heat inactivated at 68°C for 10 minutes. This procedure eliminated any *Taq* polymerase bound to the product that could inhibit 5 subsequent restriction digests. The buffer was changed to a TE buffer, and the excess PCR primers were removed with a Centricon 100 microconcentrator.

The amplified fragment was digested with *Sph*I, then treated with Klenow to create a blunt end at the *Sph*I-cleaved end, and finally digested with *Xba*I. The resulting fragment was ligated with plasmid pTma13 (pTma12-1 would have been 10 suitable) that had been digested with *Nco*I, repaired with Klenow, and then digested with *Xba*I. The ligation yielded an in-frame coding sequence with the region between the initial *Nco*I site (upstream of the first methionine codon of the coding sequence) and the introduced *Sph*I site (upstream of the methionine codon at position 140) deleted. The resulting expression vector was designated pTma16.

15 The primers used in this example are given below and in the Sequence Listing section.

<u>Primer</u>	<u>SEQ ID NO:</u>	<u>Sequence</u>
FL63	SEQ ID NO: 30	5'GATAAAGGCATGCTTCAGCTTGTGAACG
FL69	SEQ ID NO: 31	5'TGTACTTCTAGAAGCTAACAGCAG

20 Example 8

Elimination of Undesired RBS in MET140 Expression Vectors

Reduced expression of the MET140 form of *Tma* DNA polymerase can be achieved by eliminating the ribosome binding site (RBS) upstream of the methionine codon at position 140. The RBS was be eliminated via oligonucleotide site-directed 25 mutagenesis without changing the amino acid sequence. Taking advantage of the redundancy of the genetic code, one can make changes in the third position of codons to alter the nucleic acid sequence, thereby eliminating the RBS, without changing the amino acid sequence of the encoded protein.

A mutagenic primer (FL64) containing the modified sequence was synthesized 30 and phosphorylated. Single-stranded pTma09 (a full length clone having an *Nco*I site) was prepared by coinfecting with the helper phage R408, commercially available from Stratagene. A "gapped duplex" of single stranded pTma09 and the large fragment from the *Pvu*II digestion of pBS13+ was created by mixing the two plasmids, heating to boiling for 2 minutes, and cooling to 65°C for 5 minutes. The phosphorylated primer 35 was then annealed with the "gapped duplex" by mixing, heating to 80°C for 2 minutes, and then cooling slowly to room temperature. The remaining gaps were filled by extension with Klenow and the fragments ligated with T4 DNA ligase, both reactions

taking place in 200  $\mu$ M of each dNTP and 40  $\mu$ M ATP in standard salts at 37°C for 30 minutes.

5 The resulting circular fragment was transformed into DG101 host cells by plate transformations on nitrocellulose filters. Duplicate filters were made and the presence of the correct plasmid was detected by probing with a  $\gamma$ <sup>32</sup>P-phosphorylated probe (FL65). The vector that resulted was designated pTma19.

10 The RBS minus portion from pTma19 was cloned into pTma12-1 via an NcoI/XbaI fragment switch. Plasmid pTma19 was digested with NcoI and XbaI, and the 620 bp fragment was purified by gel electrophoresis, as in Example 7, above. Plasmid pTma12-1 was digested with NcoI, XbaI, and XcmI. The XcmI cleavage inactivates the RBS+ fragment for the subsequent ligation step, which is done under conditions suitable for ligating "sticky" ends (dilute ligase and 40  $\mu$ M ATP). Finally, the ligation product is transformed into DG116 host cells for expression and designated pTma19-RBS.

15 The oligonucleotide sequences used in this example are listed below and in the Sequence Listing section.

<u>Oligo</u>	<u>SEQ ID NO:</u>	<u>Sequence</u>
FL64	SEQ ID NO: 32	5'CTGAAGCATGTCTTGTACCGGTTACTATGAATAT
FL65	SEQ ID NO: 33	5'TAGTAACCGGTGACAAAG

20

### Example 9

#### Expression of Truncated Tma DNA Polymerase MET ASP21

To effect translation initiation at about the aspartic acid codon at position 21 of the Tma DNA polymerase gene coding sequence, a methionine codon is introduced before the codon, and the region from the initial NcoI site to this introduced methionine codon is deleted. The deletion process involves PCR with the same downstream primer described above (FL69) and with an upstream primer (FL66) designed to incorporate an NcoI site and a methionine codon to yield a 570 base pair product.

30 The amplified product is concentrated with a Centricon-100 microconcentrator to eliminate excess primers and buffer. The product is concentrated in a Speed Vac concentrator and then resuspended in the digestion mix. The amplified product is digested with NcoI and XbaI. Likewise, pTma12-1, pTma13, or pTma19-RBS is digested with the same two restriction enzymes, and the digested, amplified fragment is ligated with the digested expression vector. The resulting construct has a deletion from the NcoI site upstream of the start codon of the native Tma coding sequence to the new methionine codon introduced upstream of the aspartic acid codon at position 21 of the native Tma coding sequence.

35 Similarly, a deletion mutant can be created such that translation initiation begins at Glu74, the glutamic acid codon at position 74 of the native Tma coding sequence.

An upstream primer (FL67) is designed to introduce a methionine codon and an NcoI site before Glu74. The downstream primer and cloning protocol used are as described above for the MET-ASP21 construct.

The upstream primer sequences used in this example are listed below and in the 5 Sequence Listing section.

<u>Oligo</u>	<u>SEQ ID NO:</u>	<u>Sequence</u>
FL66	SEQ ID NO: 34	5'CTATGCCATGGATAGATCGCTTCTACTTCC
FL67	SEQ ID NO: 35	5'CAAGCCCATGGAAACTTACAAGGCTCAAAGA

#### Example 10

##### Expression Vectors With T7 Promoters

10 Expression efficiency can be altered by changing the promoter and/or ribosomal binding site (RBS) in an expression vector. The T7 Gene10 promoter and RBS have been used to drive expression of Tma DNA polymerase from expression vector pTma17, and the T7 Gene10 promoter and the Gene N RBS have been used to drive 15 expression of Tma DNA polymerase from expression vector pTma18. The construction of these vectors took advantage of unique restriction sites present in pTma12-1: an AfII site upstream of the promoter, an NcoI site downstream of the RBS, and a BspEI site between the promoter and the RBS. The existing promoter was 20 excised from pTma12-1 and replaced with a synthetic T7 Gene 10 promoter using techniques similar to those described in the previous examples.

The synthetic insert was created from two overlapping synthetic oligonucleotides. To create pTma17 (with T7 Gene 10 RBS), equal portions of FR414 and FR416 were mixed, heated to boiling, and cooled slowly to room temperature. The hybridized oligonucleotides were extended with Klenow to create a full length 25 double stranded insert. The extended fragment was then digested with AfII and NcoI, leaving the appropriate "sticky" ends. The insert was cloned into plasmid pTma12-1 digested with AfII and NcoI. DG116 host cells were transformed with the resulting plasmid and transformants screened for the desired plasmid.

30 The same procedure was used in the creation of pTma18 (with Gene N RBS), except that FR414 and FR418 were used, and the extended fragment was digested with AfII and BspEI. This DNA fragment was substituted for the P<sub>L</sub> promoter in plasmid pTma12-1 that had been digested with AfII and BspEI.

Plasmids pTma17 and pTma18 are used to transform E. coli host cells that have been modified to contain an inducible T7 DNA polymerase gene.

35 The oligonucleotides used in the construction of these vectors are listed below and in the Sequence Listing section.

FR414 SEQ ID NO: 36

5'TCAGCTTAAGACTTCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTT-

CCCTC

FR416 SEQ ID NO: 37

5'TCGACCATGGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGAAACC-

GTTG

FR418 SEQ ID NO: 38

5'TCAGTCCGGATAAACAAAATTATTTCTAGAGGGAAACCGTTG

Example 11

10

Translational Coupling

As described above, translational coupling can increase the efficiency of expression of a protein by coupling a short coding sequence just upstream of the initiation site of the coding sequence for the protein. Termination of translation of the upstream coding sequence leaves the ribosome in close proximity to the initiation site for the downstream coding sequence. The upstream coding sequence functions only to move the ribosome downstream to the start of the coding sequence for the desired protein.

20 Translationally coupled Tma expression vectors were constructed with the translation initiation signal and first ten codons of the T7 bacteriophage major capsid protein (gene 10) fused in-frame to the last six codons of E. coli TrpE placed upstream of the Tma coding region. The TGA (stop) codon for TrpE is "coupled" with the ATG (start) codon for the Tma gene, forming the sequence TGATG. A one base frame-shift is required between translation of the short coding sequence and translation of the Tma coding sequence.

25

In the example below, a fragment containing the T7 Gene 10-E. coli TrpE/TrpD fusion product (the last 6 codons and TGA stop codon from TrpE along with the overlapping ATG start codon from TrpD) was transferred from a pre-existing plasmid. One of ordinary skill will recognize that the T7 Gene 10-E. coli TrpE/TrpD fusion product used in the construction of the translationally coupled expression vectors can be constructed as a synthetic oligonucleotide. The sequence for the inserted fragment is listed below and in the Sequence Listing section.

30

The T7 Gene 10-E. coli TrpE/TrpD fusion product was amplified from plasmid pSYC1868 with primers FL48 and FL49. With primers FL51 and FL53, the 5' end of the Tma Pol I gene in pTma08 (a full length clone containing an NdeI site) was amplified from the ATG start codon to the MroI site downstream of the ATG start codon. The primers FL51 and FL49 were designed to leave overlapping regions such that the two amplified products could be annealed and extended, essentially as

described in Example 10. The two amplification products were mixed, heated to 95°C, slowly cooled to room temperature to anneal, and extended with Taq polymerase.

5 The extended insert was amplified with primers FL48 and FL53 and then digested with XbaI and MroI. Plasmid pTma12-1 was digested with MroI and treated with calf intestine alkaline phosphatase to prevent re-ligation. The digested pTma12-1 was ligated with the insert. DG116 host cells were transformed with the resulting construct and transformants screened for the desired plasmid DNA. The resulting vector was designated pTma20.

10 The sequences of the oligonucleotide primers and the T7 Gene 10-E. coli TrpE/TrpD fusion product (Gene 10 insert) are listed below and in the Sequence Listing section.

<u>Primers</u>	<u>SEQ ID NO:</u>	<u>Sequence</u>
FL48	SEQ ID NO: 39	5'TCCGGACTTTAAGAAGGAGATATAC
FL49	SEQ ID NO: 40	5'AATAGTCTAGCCATCAGAAAGTCTCCTGTGC
15 FL51	SEQ ID NO: 41	5'AGACTTCTGATGGCTAGACTATTCTT
FL53	SEQ ID NO: 42	5'CTGAATCAGGAGACCCGGGTCTTGGTC
Gene10 insert	SEQ ID NO:	5'CTTTAAGAAGGAGATATACATATGGCTAGCATGACTGGTGGACAGCAAATG CATGCACAGGAGACTTCTGATG

20

### Example 12

#### Arg U tRNA Expression

The pattern of codon usage differs between Thermotoga maritima and E. coli. In the Tma coding sequence, arginine is most frequently coded for by the "AGA" codon, whereas this codon is used in low frequency in E. coli host cells. The 25 corresponding "Arg U" tRNA appears in low concentrations in E. coli. The low concentration in the host cell of Arg tRNA using the "AGA" codon may limit the translation efficiency of the Tma polymerase gene. The efficiency of translation of the Tma coding sequence within an E. coli host may be improved by increasing the concentration of this tRNA species by cloning multiple copies of the tRNA gene into 30 the host cell using a second expression vector that contains the gene for the "Arg U" tRNA.

The Arg U tRNA gene was PCR amplified from E. coli genomic DNA using the primers DG284 and DG285. The amplification product was digested with Sall and BamHI. The ColE1 compatible vector pACYC184 was digested with Sall and BamHI, and the Arg U gene fragment was subsequently ligated with the digested vector. 35 DG101 cells were transformed, and the ligated vector was designated pARG01. Finally, DG116 host cells were co-transformed with pARG01 and pTma12-1.

The oligonucleotide primers used in this Example are listed below and in the Sequence Listing section.

<u>Primers</u>	<u>SEQ ID NO:</u>	<u>Sequence</u>
DG284	SEQ ID NO: 43	5'CGGGGATCCAAAAGCCATTGACTCAGCAAGG
5 DG285	SEQ ID NO: 44	5'GGGGGTCGACGCATGCGAGGAAAATAGACG

We Claim:

1. A purified thermostable DNA polymerase I enzyme that catalyzes combination of nucleoside triphosphates to form a nucleic acid strand complementary to a nucleic acid template strand, said enzyme derived from the eubacterium Thermotoga maritima.  
5
2. The enzyme of claim 1 having a molecular weight between about 97 kilodaltons and 103 kilodaltons.
- 10 3. The enzyme of claim 1 that has reverse transcriptase activity.
4. The enzyme of claim 1 that has 3'→5' exonuclease activity.
- 15 5. The enzyme of claim 1 that is in native form.
6. The enzyme of claim 2 that is in native form.
7. The enzyme of claim 1 that is in recombinant form.
- 20 8. The enzyme of claim 2 that is in recombinant form.
9. The enzyme of claim 2 that has reverse transcriptase and 3'→5' exonuclease activity.
- 25 10. The enzyme of claim 9 that is in native form.
11. The enzyme of claim 9 that is in recombinant form.
12. A method for purifying Thermotoga maritima DNA polymerase from T. maritima cells, said method comprising the steps of:  
30 (a) preparing a crude cell extract from said cells;  
(b) adjusting the ionic strength of said extract so that said polymerase dissociates from any nucleic acid in said extract;  
(c) subjecting the extract to hydrophobic interaction chromatography;  
(d) subjecting the extract to DNA binding protein affinity chromatography;  
35 (e) subjecting the extract to nucleotide binding protein affinity chromatography; and  
(f) subjecting the extract to chromatography selected from the group consisting of anion exchange, cation exchange, and hydroxyapatite chromatography.  
40

13. A recombinant DNA sequence that encodes Thermotoga maritima DNA polymerase I activity.

5 14. The DNA sequence of claim 13 that encodes the amino acid sequence of SEQ ID NO: 1, from amino to carboxy terminus:

1 MARLFLFDGT ALAYRAYYAL DRSLSTSTGI PTNATYGVAR MLVRFIKDHI  
51 IVGKDYVAVA FDKKAATFRH KLLETYKAQR PKTPDILLIQQ LPYIKKLVEA  
101 LGMKVLEVEG YEADDIIATL AVKGLPLFDE IFIVTGDKDM LQLVNEKIKV  
10 151 WRIVKGISDL ELYDAQKVKE KYGVEPQQIP DLLALTGDEI DNIPGVTGIG  
201 EKTAVQLLEK YKDLEDILNH VRELPQKVRK ALLRDRENAI LSKKLAILET  
251 NVPIEINWEE LRYQGYDREK LLPLLKELEF ASIMKELQLY EESEPVGYRI  
301 VKDLVEFEKL IEKLRESPSF AIDLETSSLD PFDCDIVGIS VSFKPKEAYY  
351 IPLHHRNAQN LDEKEVLKKL KEILEDPGAK IVGQNLKFDY KVLMVKGVEP  
15 401 VPPYFDTMIA AYLLEPNEKK FNLDLALKF LGYKMTSYQE LMSFSFPLFG  
451 FSFADVPVEK AANYSCEDAD ITYRLYKTLS LKLHEADLEN VFYKIEMLV  
501 NVLARMELNG VYVDTEFLKK LSEEYGGKLE ELAEEIYRIA GEPFNINSPK  
551 QVSRILFEKL GIKPRGKTTK TGDYSTRIEV LEELAGEHEI IPLILEYRKI  
601 QKLKSTYIDA LPKMVNPKTG RIHASFNQTG TATGRLSSSD PNQLNLPTKS  
20 651 EEGKEIRKAI VPQDPNWIV SADYSQIELR ILAHLSGDEN LLRAFEEGID  
701 VHTLTASRIF NVKPEEVTEE MRRAGKMVNFI SIIYGVTPYG LSVRLGVPVK  
751 EAEKMINVNYF VLYPKVRDYI QRVVSEAKEK GYVRTLFGRK RDIPQLMARD  
801 RNTQAEGERI AINTPIQGTA ADIIKLAMIE IDRELKERKM RSKMIIQVHD  
851 ELVFEVPNEE KDALVELVKD RMTNVVKLSV PLEVDTVIGK TWS

25

15. The DNA sequence of claim 14 that is the DNA sequence of SEQ ID NO: 1:

1 ATGGCGAGAC TATTCTCTT TGATGGAAC T GCTCTGGCCT ACAGAGCGTA  
51 CTATGCGCTC GATAGATCGC TTTCTACTTC CACCGGCATT CCCACAAACG  
30 101 CCACATACGG TGTGGCGAGG ATGCTGGTGA GATTCACTCAA AGACCATATC  
151 ATTGTCGGAA AAGACTACGT TGCTGTGGCT TTGACAAAAA AAGCTGCCAC  
201 CTTCAGACAC AAGCTCCTCG AGACTTACAA GGCTCAAAGA CCAAAGACTC  
251 CGGATCTCCT GATTCAAGCAG CTTCCGTACA TAAAGAAGCT GGTCGAAGGCC  
301 CTTGGAATGA AAGTGCTGGA GGTAGAAGGA TACGAAGCGG ACGATATAAT  
35 351 TGCCACTCTG GCTGTGAAGG GGCTTCCGCT TTTGATGAA ATATTCAAG  
401 TGACCGGAGA TAAAGACATG CTTCAGCTTG TGAACGAAAAA GATCAAGGTG  
451 TGGCGAATCG TAAAAGGGAT ATCCGATCTG GAACTTTACG ATGCGCAGAA  
501 GGTGAAGGAA AAATACGGTG TTGAACCCCA GCAGATCCCG GATCTTCTGG  
551 CTCTAACCGG AGATGAAATA GACAACATCC CCGGTGTAAC TGGGATAGGT

601 GAAAAGACTG CTGTTCAGCT TCTAGAGAAG TACAAAGACC TCGAAGACAT  
651 ACTGAATCAT GTTCGCGAAC TTCCCTAAAAA GGTGAGAAAAA GCCCTGCTTC  
701 GAGACAGAGA AAACGCCATT CTCAGCAAAA AGCTGGCGAT TCTGGAAACA  
751 AACGTTCCCA TTGAAATAAA CTGGGAAGAA CTTCGCTACC AGGGCTACGA  
5 801 CAGAGAGAAA CTCTTACAC TTTGAAAGA ACTGGAATTC GCATCCATCA  
851 TGAAGGAACT TCAACTGTAC GAAGAGTCCG AACCCGTTGG ATACAGAATA  
901 GTGAAAGACC TAGTGGATT TGAAAAACTC ATAGAGAAAC TGAGAGAATC  
951 CCCTTCGTTGCCATAGATC TTGAGACGTC TTCCCTCGAT CCTTCGACT  
1001 GCGACATTGT CGGTATCTCT GTGTCTTCA AACCAAAGGA AGCGTACTAC  
10 1051 ATACCACTCC ATCATAGAAA CGCCCAGAAC CTGGACGAAA AAGAGGTTCT  
1101 GAAAAAGCTC AAAGAAATTG TGGAGGACCC CGGAGCAAAG ATCGTTGGTC  
1151 AGAATTGAA ATTGATTAC AAGGTGTTGA TGGTGAAGGG TGGTGAACCT  
1201 GTTCCCTCCTT ACTTCGACAC GATGATAGCG GCTTACCTTC TTGAGCCGAA  
1251 CGAAAAGAAG TTCAATCTGG ACGATCTCGC ATTGAAATTT CTTGGATACA  
15 1301 AAATGACATC TTACCAAGAG CTCATGTCCT TCTCTTTCC GCTGTTGGT  
1351 TTCAGTTTG CCGATGTTCC TGTAGAAAAA GCAGCGAACT ACTCCTGTGA  
1401 AGATGCAGAC ATCACCTACA GACTTACAA GACCCTGAGC TAAACTCC  
1451 ACGAGGCAGA TCTGGAAAAC GTGTTCTACA AGATAGAAAT GCCCCTGTG  
1501 AACGTGCTTG CACGGATGGA ACTGAACGGT GTGTATGTGG ACACAGAGTT  
20 1551 CCTGAAGAAA CTCTCAGAACAG AGTACGGAAA AAAACTCGAA GAACTGGCAG  
1601 AGGAAATATA CAGGATAGCT GGAGAGCCGT TCAACATAAA CTCACCGAAG  
1651 CAGGTTCAA GGATCCTTT TGAAAAACTC GGCATAAAAC CACGTGGTAA  
1701 AACGACGAAA ACGGGAGACT ATTCAACACG CATAGAAGTC CTCGAGGAAC  
1751 TTGCCGGTGA ACACGAAATC ATTCCCTCTGA TTCTTGAATA CAGAAAGATA  
25 1801 CAGAAATTGA AATCAACCTA CATAGACGCT CTTCCAAGA TGGTCAACCC  
1851 AAAGACCGGA AGGATTGATG CTTCTTCAA TCAAACGGGG ACTGCCACTG  
1901 GAAGACTTAG CAGCAGCGAT CCCAATCTTC AGAACCTCCC GACGAAAAGT  
1951 GAAGAGGGAA AAGAAATCAG GAAAGCGATA GTTCCTCAGG ATCCAAACTG  
2001 GTGGATCGTC AGTGCCGACT ACTCCAAAT AGAACTGAGG ATCCTCGCCC  
30 2051 ATCTCAGTGG TGATGAGAAT CTTTGAGGG CATTGAAAGA GGGCATCGAC  
2101 GTCCACACTC TAACAGCTTC CAGAATATTC AACGTGAAAC CCGAAGAAGT  
2151 AACCGAAGAA ATGCGCCGCG CTGGTAAATGGTTAATTT TCCATCATAT  
2201 ACGGTGTAAC ACCTTACGGT CTGTCGTGA GGCTTGGAGT ACCTGTGAAA  
2251 GAAGCAGAAA AGATGATCGT CAACTACTTC GTCCCTCTACC CAAAGGTGCG  
35 2301 CGATTACATT CAGAGGGTCG TATCGGAAGC GAAAGAAAAA GGCTATGTTA  
2351 GAACGCTGTT TGGAAGAAAA AGAGACATAC CACAGCTCAT GGCCCGGGAC  
2401 AGGAACACAC AGGCTGAAGG AGAACGAATT GCCATAAACCA CTCCCATACA  
2451 GGGTACAGCA CGGGATATAA TAAAGCTGGC TATGATAGAA ATAGACAGGG  
2501 AACTGAAAGA AAGAAAATG AGATCGAAGA TGATCATACA GGTCCACGAC

2551 GAACTGGTTT TTGAAGTGCC CAATGAGGAA AAGGACGCGC TCGTCGAGCT  
2601 GGTGAAAGAC AGAATGACGA ATGTGGTAAA GCTTTCAGTG CCGCTCGAAG  
2651 TGGATGTAAC CATCGGCAAA ACATGGTCGT GA

5           16. A recombinant DNA vector that comprises the DNA sequence of claim 13.

10           17. A recombinant host cell transformed with the vector of claim 16.

15           18. The recombinant host cell of claim 17 that is E. coli.

20           19. The enzyme of claim 1, having a molecular weight about 86 kilodaltons, that is in recombinant form.

25           20. The enzyme of claim 1, having a molecular weight about 70 kilodaltons, that is in recombinant form.

30           21. The DNA sequence of claim 13 that encodes amino acids number 140 to 893 of SEQ ID NO:1.

35           22. The DNA sequence of claim 21 that is nucleotides number 418 to 2682 of SEQ ID NO: 1.

40           23. The DNA sequence of claim 13 that encodes amino acids number 284 to 893 of SEQ ID NO:1.

45           24. The DNA sequence of claim 23 that is nucleotides number 850 to 2682 of SEQ ID NO: 1.

# INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 91/05753

## I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all)<sup>6</sup>

According to International Patent Classification (IPC) or to both National Classification and IPC  
 Int.C1.5 C 12 N 15/54 C 12 N 9/12

## II. FIELDS SEARCHED

Minimum Documentation Searched<sup>7</sup>

Classification System	Classification Symbols
Int.C1.5	C 12 N

Documentation Searched other than Minimum Documentation  
 to the Extent that such Documents are Included in the Fields Searched<sup>6</sup>

## III. DOCUMENTS CONSIDERED TO BE RELEVANT<sup>9</sup>

Category <sup>10</sup>	Citation of Document, <sup>11</sup> with indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>
P, X	WO,A,9109944 (CETUS CORP.) 11 July 1991, see page 13, lines 28-34; page 14, line 34 - page 16, line 4 (cited in the application)	1-11, 14 , 15
P, Y	---	12, 13, 16-24
Y	Chemical Abstracts, volume 105, no. 5, 4 August 1986, (Columbus, Ohio, US) R. Huber et al.: ""Thermotoga maritima sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90 degrees C", see page 386, abstract no. 38901u, & Arch. Microbiol. 1986, 144(4), 324-33 ---	12, 13, 16-24

<sup>6</sup> Special categories of cited documents :<sup>10</sup>

- <sup>7</sup>A" document defining the general state of the art which is not considered to be of particular relevance
- <sup>7</sup>E" earlier document but published on or after the international filing date
- <sup>7</sup>L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- <sup>7</sup>O" document referring to an oral disclosure, use, exhibition or other means
- <sup>7</sup>P" document published prior to the international filing date but later than the priority date claimed

<sup>7</sup>T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

<sup>7</sup>X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

<sup>7</sup>Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

<sup>7</sup>&" document member of the same patent family

## IV. CERTIFICATION

Date of the Actual Completion of the International Search

14-11-1991

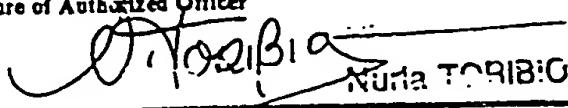
Date of Mailing of this International Search Report

20.12.91

International Searching Authority

EUROPEAN PATENT OFFICE

Signature of Authorized Officer

  
Nuria TORIBIO

III. DOCUMENTS CONSIDERED TO BE RELEVANT		(CONTINUED FROM THE SECOND SHEET)	
Category *	Citation of Document, with indication, where appropriate, of the relevant passages		Relevant to Claim No.
Y	The Journal of Biological Chemistry, volume 264, no. 11, 15 April 1989, The American Society for Biochemistry and Molecular Biology, Inc. (Baltimore, US) F.C. Lawyer et al.: "Isolation, characterization, and expression in Escherichia coli of the DNA polymerase gene from <i>thermus aquaticus</i> ", pages 6427-6437, see page 6427, left-hand column, line 1 - right-hand column. line 15 ---		12,13, 16-24
Y	EP,A,0359006 (TOYO BOSEKI KABUSHIKI KAISHA) 21 March 1990, see page 2, line 42 - page 3, line 7 -----		12,13, 16-24

ANNEX TO THE INTERNATIONAL SEARCH REPORT  
ON INTERNATIONAL PATENT APPLICATION NO.

US 9105753  
SA 51035

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 03/12/91. The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A- 9109944	11-07-91	WO-A- 9109950	11-07-91
EP-A- 0359006	21-03-90	JP-A- 2060585	01-03-90

